

**STEP Survey Weighting Procedures Summary  
(Based on The World Bank Weight Requirement)**

**Colombia**

**October 11, 2013**



## Table of Contents

1	Survey Design Overview.....	2
1.1	Target Population.....	2
1.2	Sampling.....	2
2	Data Collection Outcomes .....	5
3	Weighting Overview.....	7
4	Weighting Process Files.....	8
4.1	Revisions to Input Data File .....	8
4.2	Final Output File from Weighting Process .....	13
5	Weight Calculations based on World Bank (WB) Weight Requirement.....	17
5.1	First-stage Weight (PSU Weight).....	17
5.2	Second-stage Weight (SECTION Weight) .....	19
5.2.1	Second-stage Response Adjustment.....	20
5.2.2	Adjustment of SECTION Weights for Unknown Eligibility.....	22
5.2.3	Adjustment of SECTION Weights for Non-participation .....	22
5.2.4	Final Second-stage SECTION Weight.....	23
5.3	Third Stage: Block Weight.....	24
5.3.1	Third Stage Response Adjustment .....	25
5.3.2	Adjustment of Block Weights for Unknown Eligibility .....	26
5.3.3	Adjustment of BLOCK Weights for Non-participation .....	27
5.3.4	Third Stage Response-Adjusted BLOCK Weight.....	29
5.4	Fourth Stage: Block Segment Weight (Adjustment for Multiple Block Segments).....	29
5.4.1	Final Third/Fourth Stage BLOCK-Segment Weight.....	30
5.5	Dwelling Weight – Fifth Stage.....	31
5.5.1	Dwelling Weight - Basic.....	31
5.5.2	Dwelling Weight – Adjustment for Observed Eligibility.....	31
5.5.3	Dwelling Weight - Final .....	32
5.6	Household Weight – Sixth Stage .....	33
5.6.1	Household Weight - Basic .....	33
5.6.2	Household Weight – Adjustment for Non-Response.....	34
5.6.3	Household Weight – Adjustment for Unknown Eligibility .....	36
5.6.4	Household Weight - Adjustment for Non-participation .....	37
5.6.5	Household Weight - Final.....	37

5.7	Seventh-stage Weight (Person Weight).....	38
5.7.1	Person Weight - Basic .....	38
5.7.2	Seventh-stage Person-level Response Adjustment .....	39
5.7.3	Final Third-stage Person Weight .....	44
6	Benchmark Weights – Adjustment of Weights to Known Population Totals .....	46
6.1	Benchmark Adjusted Weight – Dwelling Level .....	46
6.2	Benchmark Adjusted Weight – Person Level .....	48

# 1 Survey Design Overview

The Colombia survey firm implemented a Full literacy assessment design. The Full assessment required each selected person to attempt to complete a General Booklet comprising Reading Components and a set of Core Literacy Items. Subsequently, those respondents that correctly answered at least 3 Core Literacy items were assigned one of four assessment Exercise Booklets to complete. The full assessment sampling objective was to have a minimum of about 2400 completed Exercise Booklets, with a distribution of about 600 complete cases for each of the four Exercise Booklets.

## 1.1 Target Population

The STEP target population comprises all non-institutionalized persons 15 to 64 years of age (inclusive) living in private dwellings in urban areas of the thirteen major Metropolitan Areas (MAs) of the country at the time of data collection; the metropolitan areas covered are thirty-three municipalities.

### Exclusions

Urban areas outside the thirteen major MAs of the country are excluded; these urban exclusions comprise approximately 35% of the urban target population.

## 1.2 Sampling

The Colombia sample design is a stratified 7 stage sample design. The stratification variable is city-size category.

The Table 1 sample frame information is excerpted from the Colombia firm's documentation of the final Colombia sampling plan, i.e., '*FINAL SAMPLING PLAN (ARD-397)-June 2013*', i.e., the Word document '[ARD\\_397\\_Final\\_sampling\\_plan\\_Colombia\\_060813](#)'; according to the sampling plan document, the table provides the sample frame counts for the Colombia target population.

**Table 1: STRUCTURE OF THE TARGET POPULATION**

Stratum	Metropolitan area (MA)	Municipalities	Dwellings	Households (B)	Population (A)	A/B	Population 15 -64	Total # of sections	Total # of blocks
1	BOGOTA	1	1,758,344	1,927,390	6,725,493	3.5	4,387,012	2,769	37,096
1	A.M MEDELLIN	10	847,403	849,992	3,116,593	3.7	2,044,986	1,326	15,366
1	A.M CALI	2	524,536	566,371	2,107,232	3.7	1,333,425	920	12,770
1	A.M B/QUILLA	2	324,770	358,658	1,560,793	4.4	959,843	830	11,820
1	A.M B/MANGA	4	229,355	252,103	944,633	3.7	605,385	501	6,560
2	A.M CUCUTA	4	169,163	176,353	705,185	4.0	422,460	519	7,198
2	CARTAGENA	1	183,255	195,056	842,581	4.3	544,336	455	6,083
2	PASTO	1	76,888	80,307	308,665	3.8	205,231	153	2,318
2	IBAGUE	1	121,636	128,661	461,739	3.6	285,455	251	3,682
2	A.M PEREIRA	3	163,191	167,164	602,678	3.6	347,238	323	5,107
2	A.M MANIZALES	2	106,522	105,631	373,698	3.5	249,308	243	3,174
2	MONTERIA	1	61,347	64,536	286,229	4.4	180,092	195	3,152
2	VILLAVICENCIO	1	90,133	96,907	354,551	3.7	215,640	262	3,690
	<b>13 Areas</b>	<b>33</b>	<b>4,656,543</b>	<b>4,969,129</b>	<b>18,390,070</b>	<b>3.7</b>	<b>11,780,411</b>	<b>8,747</b>	<b>118,016</b>

Source: SEI s.a. based on aggregates per block from 2005 Population Census Dane, basic household form.

### First Stage Sample

The primary sample unit (PSU) is a metropolitan area. A sample of 9 metropolitan areas was selected from the 13 metropolitan areas on the sample frame.

The metropolitan areas were grouped according to city-size; the five largest metropolitan areas are included in Stratum 1 and the remaining 8 metropolitan areas are included in Stratum 2. The five metropolitan areas in Stratum 1 were selected with certainty; in Stratum 2, four metropolitan areas were selected with probability proportional to size (PPS), where the measure of size was the number of persons aged 15 to 64 in a metropolitan area.

### Second Stage Sample

The second stage sample unit is a Section. At the second stage of sample selection, a PPS sample of 267 Sections was selected from the sampled metropolitan areas; the measure of size was the number of persons aged 15 to 64 in a Section. The sample of 267 Sections consisted of 243 initial Sections and 24 reserve Sections to be used in the event of complete non-response at the Section level.

### Third Stage Sample

The third stage sample unit is a Block. Within each selected Section, a PPS sample of 4 blocks was selected; the measure of size was the number of persons aged 15 to 64 in a Block. Two sample Blocks were initially activated while the remaining two sample Blocks were reserved for use in cases where there was a refusal to cooperate at the Block level or cases where the block did not belong to the target population (e.g., parks, and commercial and industrial areas).

### Fourth Stage Sample

The fourth stage sample unit is a Block Segment.

Regarding the Block segmentation strategy, the Colombia document 'FINAL SAMPLING PLAN (ARD-397)' states *"According to the 2005 population and housing census conducted by DANE, the average number of dwellings per block in the 13 large cities or metropolitan areas was approximately 42 dwellings. Based on this finding, the defined protocol was to report those cases in which 80 or more dwellings were present in a given block in order to partition block using a random selection algorithm."*

At the fourth stage of sample selection, 1 Block Segment was selected in each selected Block using a simple random sample (SRS) method.

### Fifth Stage Sample

The fifth stage sample unit is a dwelling. At the fifth stage of sample selection, 5582 dwellings were selected from the sampled Blocks/Block Segments using a simple random sample (SRS) method.

According to the Colombia document 'FINAL SAMPLING PLAN (ARD-397)', the selection of dwellings within a participant Block *"was performed differentially amongst the different socioeconomic strata that the Colombian government uses for the generation of cross-subsidies for public utilities (in this case, the socioeconomic stratum used for the electricity bill was used). Given that it is known from previous survey implementations that refusal rates are highest amongst households of higher socioeconomic status, the number of dwellings to be selected increased with the socioeconomic stratum (1 being the poorest and 6 being the richest) that was most prevalent in a given block".*<sup>1</sup>

---

<sup>1</sup> The prevalent socioeconomic category is used as a criterion to decide on the number of dwellings to select within a sampled Block; refer to Table 2.

The following table from the document 'FINAL SAMPLING PLAN (ARD-397)' shows the number of dwellings selected according to the prevalent socioeconomic stratum within the participant Block.

<b>Table 2:</b> Number of dwellings selected according to the prevalent socioeconomic stratum within the Block	
<b>Prevalent socioeconomic stratum in the block</b>	<b>Number of dwellings selected</b>
1	8
2	9
3	11
4	13
5	15
6	15

Source: SEI s. a. calculations

#### Sixth Stage Sample

The sixth stage sample unit is a household. At the sixth stage of sample selection, one household was selected in each selected dwelling using an SRS method.

#### Seventh Stage Sample

The seventh stage sample unit was an individual aged 15-64 (inclusive). The sampling objective was to select one individual with equal probability from each selected household.

Table 3 provides the counts of selected Sections, selected Blocks, selected Dwellings, selected Households, and selected Persons by sampled Stratum, and by Metropolitan Area.

**Table 3: Colombia Sample Counts [Sectors, Blocks, Dwellings, Households (HH), Persons]**

<b>Stratum</b>	<b>Metro Area</b>	<b>Sections</b>	<b>Blocks</b>	<b>Selected Dwellings</b>	<b>HHs in Selected Dwellings</b>	<b>Selected HHs</b>	<b>Selected Persons</b>
<b>1 (Certainty inclusion)</b>	<b>Medellín</b>	44	176	967	1014	967	441
	<b>Barranquilla</b>	23	92	349	413	347	243
	<b>Bogotá</b>	93	364	2,304	2602	2287	1026
	<b>Bucaramanga</b>	16	63	291	314	283	169
	<b>Cali</b>	30	118	546	602	545	319
<b>Sub-Total Stratum 1</b>		<b>206</b>	<b>813</b>	<b>4457</b>	<b>4945</b>	<b>4429</b>	<b>2198</b>
<b>2 (Probability inclusion)</b>	<b>Manizales</b>	13	52	235	256	233	131
	<b>Villavicencio</b>	12	48	197	210	197	112
	<b>Cúcuta</b>	21	84	454	467	442	225
	<b>Ibagué</b>	16	64	239	265	239	158
<b>Sub-Total Stratum 1</b>		<b>62</b>	<b>248</b>	<b>1125</b>	<b>1198</b>	<b>1111</b>	<b>626</b>
<b>Grand Total</b>		<b>267</b>	<b>1,061</b>	<b>5,582</b>	<b>6,143</b>	<b>5,540</b>	<b>2824</b>

## 2 Data Collection Outcomes

The final survey weights depend on the outcome from the attempt to obtain an interview at each selected household. The treatment of each sampled case in the weighting process depends on the final 'Result Code of Interview' assigned to the sampled household and the final 'Result Code of Interview' assigned to the selected person. The result codes used in the Colombia STEP are listed in Table 4; a 'Result Code of Interview' was assigned to each sampled case.

<b>Table 4: <i>Full Literacy Assessment</i> Final Result Codes for Sampled Households and Selected Persons</b>	
<b>i) Household-level Codes</b>	
<b>Code</b>	<b>Code Description</b>
01	Household Refusal
02	Household unusual circumstance (e.g., death in family, illness, fire in dwelling, etc.)
03	No knowledgeable household member found after 3 visits (only a child, non-competent adult, etc.)
04	Temporarily absent/unavailable for field period (information from other persons)
05	No competent household member to interview (because of severe illness, mental disability, etc.)
06	Language problem – no household member spoke a language understandable by survey team, no translator available
07	Dwelling could not be found / selected address has no household
08	Vacant dwelling
09	Dwelling not habitable/dwelling destroyed/dwelling converted to commercial use
10	No answer at door
21	Household Refusal to continue Household Module (i.e., begun but not complete)
22	Household unusual circumstance (e.g., death in family, illness, fire in dwelling, etc.) prevented continuance
31	No household member in the eligible range of 15-64
94	Case dropped due to data quality issue
95	Sample case justifiably not activated, e.g., reserve sample household not needed, sample goal achieved
96	General non-response; Reason unspecified by survey firm
97	Household not interviewed due to field procedure error – wrong household selected in its place
<b>j) Person-level Codes</b>	
<b>Code</b>	<b>Code Description</b>
	<b>Individual modules 2-7 not begun or unacceptable interview</b>
32	Refusal by Selected Person to begin modules 2-7
33	Unusual circumstance - Selected Person (e.g., death in family, illness, fire in dwelling, etc.)
34	Refusal - Another household member refused to let selected individual do individual modules
35	Unable to contact Selected Person after three visits during field period
36	Temporarily absence/unavailability of Selected Person for field period (info from household member)
37	Hearing Problem - Selected Person is deaf or so hard-of-hearing, no translator available
38	Learning or mental disability (includes emotional conditions like severe depression) of Selected Person
39	Speech impairment of Selected Person prevented participation, no translator available
40	Language problem – Selected Person's language not understood by survey team, no translator available
98	Sample selection error – an incorrect eligible person was selected
99	Sample selection error – an ineligible person was selected
	<b>Individual modules 2-7 begun but not completed</b>
51	Refusal by Selected Person to continue modules 2-7
52	Unusual circumstance - Selected Person
	<b>General Booklet (module 9) not begun</b>
61	Unable to read or write in language of General Booklet – Selected Person refused to begin
62	Refusal by Selected Person to begin General Booklet
63	Unusual circumstance - Selected Person
64	Blindness or visual impairment of Selected Person prevented completion of General Booklet
65	Physical disability of Selected Person prevented completion of General Booklet



<b>Table 4: <i>Full Literacy Assessment</i> Final Result Codes for Sampled Households and Selected Persons</b>	
	<b>General Booklet (module 9) looked through or attempted but not completed and not passed</b>
71	Selected Person looked at Booklet but unable to read & write the language of Booklet - did not begin
72	Refusal by Selected Person to continue the General Booklet
73	Interruption – unacceptably long
74	Unusual circumstance - Selected Person
75	Attempted General Booklet but no answers marked
76	Attempted part or all of General Booklet; less than 3 correct Core answers
	<b>Exercise Booklet (module 9)</b>
81	No attempt at any of the exercise booklet items / refusal to start
82	Interruption too long during exercise booklet administration
83	Unusual circumstance, Exercise Booklet administration terminated
84	Attempt at some Exercise Booklet items but no answers marked
85	Attempt at all Exercise Booklet items but no answers marked
86	Marked some Exercise Booklet items, not all items attempted
87	Marked some Exercise Booklet items, all items attempted
88	Marked all Exercise Booklet items

### 3 Weighting Overview

The Colombia stratified seven-stage cluster design resulted in differential probabilities of selection for the selected persons. Consequently, each selected person in the survey does not necessarily represent the same number of persons in the target population. To account for differential probabilities of selection due to the nature of the design and to ensure accurate survey estimates, the Colombia STEP requires a sampling weight for each person that participated in the survey.

In general, the objectives of the Colombia STEP weighting are to construct a set of survey weights to,

- 1) compensate for unequal probabilities of selection;
- 2) compensate for household-level non-response and person-level non-response;
- 3) adjust the weighted sample distribution of the number of dwellings by stratum and the number of persons by gender and by stratum so that the weighted survey estimates for these variables conform to a known population distribution for these variables.

The general weighting procedure for the Colombia STEP survey required the following tasks.

- 1) Creation of a data file to input into the weighting process;

- 2) **World Bank (WB) Weight Requirement:**

Create survey weights for sampled cases of households and persons that provided sufficient data to be considered a participant in the survey. This requirement does not necessarily include the completion of an assessment General Booklet, and an Exercise Booklet, and all household and individual questionnaire modules.

- a) Calculation of a PSU weight for 9 activated sample PSUs;
  - b) Calculation of a household weight for each sampled household;
    - i) Calculation of a household-level non-response adjustment independently for each PSU;
    - ii) Benchmark adjustment of the dwelling & household weights to known population totals for the number of dwellings.
  - c) Calculation of a person weight for each selected person (SP);
    - i) Calculation of a non-response adjustment independently for each sampled person.
    - ii) Benchmark adjustment of the person weights to known population totals for age and gender.
- 3) The required output from the weighting process is a final Colombia data file with the survey design weights (i.e., for each sampled PSU, household, person) and benchmark weights appended to each data record.

## 4 Weighting Process Files

The key file for the weighting process is the file '[archivo\\_factores\\_010813\\_v10.por](#)' provided by the WB core team. The file '[archivo\\_factores\\_010813\\_v10.por](#)' contains a data record for the 22,495 listed households in the selected cooperating Blocks; this file includes 5582 sampled households. An SPSS version of this file, i.e., '[archivo\\_factores\\_010813\\_v10.sav](#)' was created for the weighting process.

### 4.1 Revisions to Input Data File

The following general revisions were made to the Colombia SPSS data file '[archivo\\_factores\\_010813\\_v10](#)':

- 1) '*m11\_q07*', the Result of Interview Code: revised the value labels to correspond to the Codes and code descriptions in Table 3 of this document.
- 2) '*ma*', Metropolitan Area: revised value labels
  - a) 1- Bogotá
  - b) 2- Medellín A.M.
  - c) 3-Cali A.M.
  - d) 4-Barranquilla A.M.
  - e) 5-Bucaramanga A.M.
  - f) 7-Manizales A.M.
  - g) 9-Villavicencio
  - h) 11- Cúcuta
  - i) 13- Ibagué
- 3) New Variables needed for weighting calculations
  - a) **SectBlk\_Status**: Final status code for sampled Sections and sampled BLOCKs
    - i) Created by recoding the variable '*section*';
    - ii) Value Labels
      - ♦ 1-Complete survey
      - ♦ 2-Block out of universe
      - ♦ 3-Block not activated
      - ♦ 4-Block Refusal
      - ♦ 5-Section Refusal
      - ♦ 6-Section out of universe
      - ♦ 7-Section not activated
  - b) **Stratum\_MA**: Stratum identification for Metropolitan Area '*ma*'
    - i) Created as follows:
      - ♦ Stratum\_MA=1 if ( 1 <= ma <= 5 )
      - ♦ Stratum\_MA=2 if ( 5 <= ma <= 13 )
    - ii) Value Labels
      - ♦ 1-Certainty MAs
      - ♦ 2-Probabilistic MAs
  - c) **SBDHP\_ID**: Sequential case identification number
    - i) The Colombia final sample file was sorted in ascending order of the variables Stratum\_MA, ma, district, subdistr, sector, seccion, manzana, dwelling; subsequently '**SBDHP\_ID**' was assigned the sequential numbers 0001 to 6085.

5) Missing Values: The missing values for essential weighting variables were resolved as follows:

- a) missing values of '*nm\_ii*' (i.e., Number of Sections Selected in Second Stage)
  - i) **N\_SECTS\_Sel**: Number of sampled Sections in the sampled Metropolitan Area
    - ♦ Used in the Section weight calculations instead of the variable '*nm\_ii*' since there are missing values of '*nm\_ii*'.
    - ♦ Created as a count of the unique occurrences of the variable '*seccion*' (i.e., DANE Code for Section) within each sampled '*ma*'.
- b) missing values of '*nm\_iii0*' (i.e., Number of Selected Blocks in the Selected Section)
  - i) **N\_BLKs\_Sel**: Number of sampled Blocks in the sampled Section
    - ♦ Used in the Block weight calculations instead of the variable '*nm\_iii0*' since there are missing values of '*nm\_iii0*'.
    - ♦ Created as a count of the unique occurrences of the variable '*manzana*' (i.e., DANE Code for Block) within each sampled '*seccion*'.
- c) missing values of '*n\_viv*' (i.e., Number of Selected Blocks in Third Stage)
  - i) **DwgsLstd\_Bl**: Number of listed dwellings in the sampled Block
    - ♦ Used in the Dwelling weight calculations instead of the variable '*n\_viv*' since there are missing values of '*n\_viv*'.
      - for data records with non-missing variable *n\_viv*,
        - **DwgsLstd\_Bl** = *n\_viv*
      - for data records with missing variable *n\_viv*,
        - **DwgsLstd\_Bl** = the previous non-missing value of *n\_viv* within the sampled Block
- d) missing values of '*nm\_viv*' (i.e., Number of Selected Dwellings in The Block)
  - i) **DwgsSel\_Bl**: Number of selected dwellings in the sampled Block
    - ♦ Used in the Dwelling weight calculations instead of the variable '*nm\_viv*' since there are missing values of '*nm\_viv*'.
      - for data records with non-missing variable *nm\_viv*,
        - **DwgsSel\_Bl** = *nm\_viv*
      - for data records with missing variable *nm\_viv*,
        - **DwgsSel\_Bl** = the previous non-missing value of *nm\_viv* within the sampled Block
- e) missing values of '*n\_hog*' (i.e., Number of Households in Selected Dwellings)
  - i) **DwgHHs**: Number of Households in the selected Dwellings
    - ♦ Used in the Household weight calculations instead of the variable '*n\_hog*' since there are missing values of '*n\_hog*'.
      - for data records with non-missing variable *n\_hog*,
        - **DwgHHs** = *n\_hog*
      - for data records with missing variable *n\_hog*,
        - **DwgHHs** = *n\_hog\_median*, the Stratum median value for non-missing cases of the variable *n\_hog*
- f) missing values of '*nm\_hog*' (i.e., Number of Selected Households in Selected Dwellings)
  - i) **DwgHHs\_SEL**: Number of selected Households in the sampled Dwelling
    - ♦ Used in the Household weight calculations instead of the variable '*nm\_hog*' since there are missing values of '*nm\_hog*'.

- for data records with non-missing variable *nm\_hog*,
    - **DwgHHs\_SEL**= *n\_hog*
  - for data records with missing variable *nm\_hog*,
    - **DwgHHs\_SEL**= 1 for missing cases of the variable *nm\_hog*
- g) missing values of the variable '*m1c\_a*' (i.e., NUMBER A: Eligible Persons)
- i) **EligPersCnt**: Number of selected Persons in the sampled Household
- ♦ Used in the Person weight calculations instead of the variable '*m1c\_a*' since there are missing values of '*m1c\_a*'.
    - for data records with non-missing variable *nm\_hog*,
      - **EligPersCnt**= *m1c\_a*
    - for data records with missing variable *nm\_hog*,
  - ♦ **EligPersCnt**= *m1c\_a\_median*, the Stratum median value for non-missing cases of the variable *m1c\_a*
- 6) Selected data records for inclusion in the final sample file
- a) The data records for all sampled Sections using the variables 'selected' and 'SectBlk\_Status' were selected according to the following criterion: 'selected' = 1 or 'SectBlk\_Status' ≠ 1.
- b) The resulting final Colombia sample file consisted of 6085 data records.
- 7) **SPSS Output File**: '[W01-COL Final Data File\(02-Aug-2013\)-selected](#)'
- d) This file includes all 6085 sampled cases; each case includes a final result code for the sampled Section and the sampled Block.
- i) Table 5 provides the breakdown of the 6085 cases by the final result code for the Section/Block.
- ♦ Note that some Sections and/or Blocks were non-participants in the Colombia STEP survey.

**Table 5: Number of STEP Cases by Section/Block Result Code**

Section-Block Result Code	Stratum-Metro Area		Total
	1-Certainty MAs	2-Probabilistic MAs	
1-Complete survey	4457	1125	5582
2-Block out of universe	7	3	10
3-Block not activated	198	96	294
4-Block Refusal	58	4	62
5-Section Refusal	85	4	89
6-Section out of universe	8		8
7-Section not activated	12	28	40
<b>Total</b>	<b>4825</b>	<b>1260</b>	<b>6085</b>

- ii) For the participant Sections and Blocks (i.e., 1-Complete Survey in above Table 4) there are 5582 sampled Dwellings; the data record for each selected Dwelling includes the variable 'm11\_q07', the result code of the interview. Table 6 provides the breakdown of the 5582 dwelling cases by the final result code of the interview.

**Table 6: Colombia STEP: Number of Sampled Dwellings by Result Code of Interview 'm1a\_q07'**

<b>RESULT CODE OF INTERVIEW</b>	<b># of Dwellings</b>
01-Refusal-Household	846
02-Unusual circumstance-Household	50
03-No knowledgeable household member found after 3 visits	200
04-Temporarily absent/unavailable for field period	56
05-No competent household member (e.g., severe illness, mental disability, etc.)	16
06-Language problem	14
08-Vacant dwelling	33
09-Dwelling not habitable/destroyed/converted to commercial	9
10-No answer at door	1419
21-Household Refusal to continue Household Module (begun,not complete)	29
22-Household module terminated - unusual circumstance	50
31-No eligible household member aged 15-64	36
32-Refusal by SP to begin modules 2-7	87
33-Interview termination-SP unusual circumstance	6
34-Refusal-Other Household Member refused to let SP participate	7
35-Unable to contact SP after 3 attempts	90
36-SP absent for entire survey period	16
38-SP learning or mental disability	1
51-Refusal by SP to continue modules 2-7	17
52-Interview termination-SP unusual circumstance	4
61-SP unable to read language of GB	9
62-SP refusal to begin GB	5
64-SP blindness or visual impairment-unable to do GB	7
65-SP physical disability	3
71-SP looked through GB-unable to read/write language of GB	9
72-SP refused to continue GB after starting	3
73-GB interruption too long	1
74-GB terminated - unusual circumstance	2
75-Attempted General Booklet, no answers marked	2
76-Attempted part or all of General Booklet; <3 correct Core answers	133
81-No attempt at any exercise booklet items / refusal to start	1
82-Interruption too long during exercise booklet administration	1
84-Attempt at some Exercise Booklet items, no answers marked	2
85-Attempt at all Exercise Booklet items, no answers marked	4
86-Marked some Exercise Booklet items, not all items attempted	6
87-Marked some Exercise Booklet items, all items attempted	849
88-Marked all Exercise Booklet items	1559
<b>TOTAL</b>	<b>5582</b>

- iii) This file is the primary file for the process of creating the survey weights.
- ♦ Table 7 provides the list of key variables for weighting that are included in ‘W01-COL Final Data File(02-Aug-2013)-selected’

Table 7: Key Variables in Primary Input File for Weighting Process		
Variable Name	Source	Description
country	Colombia Data File	Country abbreviation code
estr_mu	Colombia Data File	SURVEY STRATA
ma	Colombia Data File	METROPOLITAN AREA
district	Colombia Data File	DANE CODE FOR DEPARTMENT
subdistr	Colombia Data File	DANE CODE FOR MUNICIPALITY
sector	Colombia Data File	DANE CODE FOR SECTOR
seccion	Colombia Data File	DANE CODE FOR SECTION
manzana	Colombia Data File	DANE CODE FOR BLOCK
ea	Colombia Data File	DEPARTMENT CODE + MUNICIPALITY CODE+ SECTOR+ SECTION+BLOCK
section	Colombia Data File	SECTION AND BLOCKS RESULTS
dwelling	Colombia Data File	DWELLING ID
selected	Colombia Data File	SELECTED DWELLING
n_i	Colombia Data File	NUMBER OF METROPOLITAN AREAS WITHIN STRATUM
p15_64_i	Colombia Data File	PEOPLE BETWEEN 15 -64 PER STRATUM
p15_64_0	Colombia Data File	PEOPLE BETWEEN 15 -64 PER METROPOLITAN AREA SELECTED
nm_i	Colombia Data File	NUMBER OF METROPOLITAN AREAS SELECTED WITHIN STRATUM
n_ii	Colombia Data File	NUMBER OF SECTIONS WITHIN METROPOLITAN AREA
p15_64_1	Colombia Data File	PEOPLE BETWEEN 15 -64 PER SECTION
nm_ii	Colombia Data File	NUMBER OF SECTIONS SELECTED IN SECOND STAGE
n_iii0	Colombia Data File	NUMBER OF BLOCKS WITHIN SECTIONS
p15_64_2	Colombia Data File	PEOPLE BETWEEN 15 -64 PER BLOCK
nm_iii0	Colombia Data File	NUMBER OF BLOCKS IN THIRD STAGE
n_viv	Colombia Data File	NUMBER OF DWELLINGS IN THE BLOCK
nm_viv	Colombia Data File	NUMBER OF DWELLINGS SELECTED IN THE BLOCK
hbn	Colombia Data File	DWELLING ID WITHIN BLOCK
m1c_a	Colombia Data File	NUMBER A: Eligible Persons
sexo	Colombia Data File	SEX OF SELECTED PERSON
estrato	Colombia Data File	STRATUM OF ELECTRICITY SERVICE OF DWELLING OF SELECTED PERSON
cluster	Colombia Data File	INTERNAL ID BLOCK + DWELLING ID
m1a_q04	Colombia Data File	AGE OF SELECTED PERSON
m11_q07	Colombia Data File	RESULT CODE OF INTERVIEW
n_hog	Colombia Data File	NUMBER OF HOUSEHOLDS IN SELECTED DWELLINGS
nm_hog	Colombia Data File	NUMBER OF SELECTED HOUSEHOLDS IN THE DWELLING
SectBlk_Status	Created during file preparations	Section-Block Result Code
N_SECTS_Sel	Created during file preparations	SECTIONS - # selected
N_BLKs_Sel	Created during file preparations	BLOCKS - # selected
Stratum_MA	Created during file preparations	Stratum-Metro Area

## 4.2 Final Output File from Weighting Process

The output file from the weighting process is,

### 1) COL-W01-STEP(HH) Weighted Data-Design and BM Weights (16-Sep-2013)

- a) This file contains the original data variables from the Colombia data file 'archivo\_factores\_010813\_v10' and the applicable PSU/Section/Block weights, dwelling/household weights, person weights, and other variables that were created during the weighting process.
- b) The file contains 6085 data records, of which there are 2824 participating households with a final household weight. The breakdown of the count of participating households by stratum is shown below in Table 8.

Table 8: Colombia STEP - Number of Sampled Households by Stratum		
Stratum	Metropolitan Area	Total
1-Certainty MAs	1-Bogota	2,287
	2-Medellin A.M.	967
	3-Cali A.M.	545
	4-Barranquilla A.M.	347
	5-Bucaramanga A.M.	283
Sub-Total (Certainty MAs)		4,429
2-Probabilistic MAs	7-Manizales A.M.	233
	9-Villavicencio	197
	11-Cucuta.	442
	13-Ibague	239
Sub-Total (Probabilistic MAs)		1,111
	1-Bogota	2,287
	2-Medellin A.M.	967
	3-Cali A.M.	545
	4-Barranquilla A.M.	347
	5-Bucaramanga A.M.	283
	7-Manizales A.M.	233
	9-Villavicencio	197
	11-Cucuta.	442
	13-Ibague	239
Grand Total		5,540

- c) In addition to the Colombia data file variables, the weighted data file includes additional variables that were created during the weighting process. The essential variables that are needed for estimation of survey results are included in the final weighted data file.
  - i) Table 9 provides the list of variables, created or revised during the weighting process, which are included in the final weighted data file.
  - ii) Table 10 provides the list other available variables from the weighting process that are not included in the final weighted data file; these variables are available on demand.



- iii) Table 11 provides the frequency count of result codes for participating households in the final weighted data file '[COL-W01-STEP\(HH\) Weighted Data-Design and BM Weights \(16-Sep-2013\)](#)'.

<b>Table 9: Variables Created or Revised During Weighting Process - Added to Colombia Data File</b>		
<b>Variable Name</b>	<b>Description</b>	<b>Values</b>
<b>SectBlk_Status</b>	Section Result Code	1-Complete Survey 2-Block out of universe 3-Block not activated 4-Block refusal 5-Section Refusal 6-Section out of universe 7-Section not activated
<b>Stratum_MA</b>	Stratum-Metro Area	1-Certainty Metro Areas 2-Probabilistic Metro Areas
<b>m1a_q02</b>	Respondent Gender	n/a
<b>W1_MAwT</b>	Metro Area Weight	n/a
<b>W2_Fin_SECTwt</b>	Final MA-SECTION Weight	n/a
<b>W4_BLK_Segwt</b>	BLOCK Weight; Adjusted for Multiple Partitions	n/a
<b>W4_BLKwt_Fin</b>	FINAL MA-SECTION-BLOCK Weight	n/a
<b>RespCat_Dw</b>	Dwelling Final Response Category	1-Participant Dwelling 2-Ineligible Dwelling
<b>W5_Dw_Fin</b>	Final MA-SECTION-BLOCK-DWELLING Weight	n/a
<b>RespCat_HH</b>	Final Response Category-Households	1-Participant HH 2-Ineligible HH 3-Non-participant Eligible HH 4-Non-participant Unknown Eligibility HH
<b>W6_HHwt_Fin</b>	Final MA-SECTION-BLOCK DWELLING-HH Weight	n/a
<b>W7_FinSPwt</b>	Person Weight: Final MA-SECTION-BLOCK DWELLING-HH-SP Weight	n/a
<b>BM_Dwgs</b>	Benchmark Total number of Dwellings	n/a
<b>BMwt_Dw</b>	BM Dwelling Weight, Adjusted for Benchmark Dwelling Totals	n/a
<b>BMwt_HH</b>	BM Household Weight, Adjusted for Benchmark Dwelling Totals	n/a
<b>W7_SP_BM1</b>	Person Weight, Adjusted for Non-response & Benchmark Dwelling Totals	n/a
<b>BM_Popn</b>	BM Gender Count	n/a
<b>RespCat_SP</b>	Final Response category_Selected Person	1-Participant SP 2-Non-participant, Eligible
<b>BMwt_SP</b>	BM PERSON Weight, Adjusted for Benchmark Dwellings & Population Gender Counts	n/a

<b>Table 10:</b>	<b>Table 11: Variables Created During Weighting Process (Not Added to Colombia Final Data File)</b>	
<b>Variable Name</b>	<b>Description</b>	<b>Values</b>
N_SECTS_Sel	SECTIONS - # selected	n/a
N_BLKs_Sel	BLOCKS - # selected	n/a
SECT_k	Participant SECTION - MA Cases	n/a
SECT_x	Ineligible SECTION - MA Cases	n/a
SECT_z	Non-participant; Eligible SECTION - MA Cases	n/a
SECT_u	Non-participant Unknown Eligible SECTION - MA Cases	n/a
W2_SECT	SECTION Basic Weight within MA	n/a
RespCat_SECT	SECTION-Final Response Category	1-Participant Section 2-Ineligible Section 3-Non-participant; Eligible Section 4-Non-activated Section; Unknown Eligibility
Elig_Prop_SECT	Response adjustment ratio for SECTIONS	n/a
R_adj_SECT	RESPONSE_ADJUSTED SECTION Weight	n/a
W2_SECT_Radj	# of people in Blocks selected with certainty	n/a
W3_BLK1	BLOCK Basic Weight within Section	n/a
RespCat_BlK	Block-Final Response Category (revised for unknown eligibility Blocks)	1-Participant Block 2-Ineligible Block 3-Non-participant; Eligible Block
W3_BLK_Radj	BLOCK Weight Adjusted for Non-response	n/a
W4_BLKseg	Basic Block Partition Weight	n/a
sum_Pers_CertBlks	Count of Certainty Blocks in Section	n/a
Blk_SelCertain#	Block Selection Certainty Indicator	n/a
BLK_Certain	Participant Blocks	n/a
BLK_k	Ineligible Blocks	n/a
BLK_x	Non-participant; Eligible Blocks	n/a
BLK_z	Non-participant Unknown Eligible Blocks	n/a
BLK_u	Response adjustment ratio for SECTIONS	n/a
Peop_smpBlks	Target People in Sampled Blocks	n/a
Peop_R_Blks	Target People in Responding Blocks	n/a
Peop_NonR_Blks	Target People in Non-Responding Blocks	n/a
Peop_X_Blks	Target People in Ineligible Blocks	n/a
R_adj_BLK	Response adjustment ratio for BLOCKS	n/a
DwgsLstd_BlK	Dwellings Listed in Block/Segment	n/a
DwgsSel_BlK	Dwellings Selected in Block/Segment	n/a
n_hog_median	Median # Eligible Households in Stratum	n/a
DwgHHs	# of HHs in Selected Dwelling	n/a
DwgHHs_SEL	# of Selected HHs in Selected Dwelling	n/a
W5_Dw	Dwelling Weight within selected BLOCK	n/a
HHblk_k	Participants: Weighted HH - Block	n/a
HHblk_x	Ineligibles: Weighted HH - Block	n/a
HHblk_z	Non-participant Eligibles: Weighted HH - Block	n/a
HHblk_u	Non-participant Unknown Eligibles: Weighted HH - Block	n/a
W6_HH	HH Weight- Adjusted for multiple households	n/a
EligProp_HH	Eligible HHs: Wtd Proportion of cases with known eligibility	n/a
R_adj_HH	Response adjustment ratio for HHs	n/a
W6_HH_Radj	Household Weight: Adjusted for Non-response	n/a
m1c_a_median	Median # Eligible Persons in Stratum	n/a
EligPersCnt	Eligible Person Count in Participant HH	n/a
W7_SP	Basic Person Weight within household	n/a

<b>Table 10:</b>	<b>Table 11: Variables Created During Weighting Process (Not Added to Colombia Final Data File)</b>	
<b>Variable Name</b>	<b>Description</b>	<b>Values</b>
EP_Blkl	Eligible Persons in Stratum	n/a
PP_Blkl	Participants: Weighted PERSONs - Stratum	n/a
NP_Blkl	Non-participant Eligibles: Weighted PERSONs - Stratum	n/a
R_adj_SP	Response Adjustment - Basic Person Weight	n/a
W7_SP_Radj	Person Weight: Adjusted for Non-response	n/a
DW#_Est_estr	Stratum Dwelling Est.: Wtd Sum of DWELLING Weights by Stratum	n/a
PopnEst_estr	Stratum Population: Estimated # of Persons by Stratum & Gender	n/a
AgeGrp	Age Group-5 year	n/a

**Table 11: Case Summaries: Result Code of Interview for Participant Households**

<b>RESULT CODE OF INTERVIEW</b>	<b># of Cases</b>	<b>% of Total Cases</b>
32-Refusal by SP to begin modules 2-7	87	3.1%
33-Interview termination-SP unusual circumstance	6	.2%
34-Refusal-Other Household Member refused to let SP participate	7	.2%
35-Unable to contact SP after 3 attempts	90	3.2%
36-SP absent for entire survey period	16	.6%
38-SP learning or mental disability	1	.0%
51-Refusal by SP to continue modules 2-7	17	.6%
52-Interview termination-SP unusual circumstance	4	.1%
61-SP unable to read language of GB	9	.3%
62-SP refusal to begin GB	5	.2%
64-SP blindness or visual impairment-unable to do GB	7	.2%
65-SP physical disability	3	.1%
71-SP looked through GB-unable to read/write language of GB	9	.3%
72-SP refused to continue GB after starting	3	.1%
73-GB interruption too long	1	.0%
74-GB terminated - unusual circumstance	2	.1%
75-Attempted General Booklet, no answers marked	2	.1%
76-Attempted part or all of General Booklet; <3 correct Core answers	133	4.7%
81-No attempt at any exercise booklet items / refusal to start	1	.0%
82-Interruption too long during exercise booklet administration	1	.0%
84-Attempt at some Exercise Booklet items, no answers marked	2	.1%
85-Attempt at all Exercise Booklet items, no answers marked	4	.1%
86-Marked some Exercise Booklet items, not all items attempted	6	.2%
87-Marked some Exercise Booklet items, all items attempted	849	30.1%
88-Marked all Exercise Booklet items	1559	55.2%
<b>Total</b>	<b>2824</b>	<b>100.0%</b>

## 5 Weight Calculations based on World Bank (WB) Weight Requirement

The input file for the process of calculating the weights based on the World Bank (WB) weight requirement is '[W01-COL Final Data File\(02-Aug-2013\)-selected](#)', which was created from the WB provided file '[archivo\\_factores\\_010813\\_v10.por](#)'.

### 5.1 First-stage Weight (PSU Weight)

The basic first stage weight, i.e., the PSU weight, is calculated separately for each stratum. The stratum PSU weight is based on an overall sample size of 9 urban PSUs, i.e., 9 Metropolitan Areas.

#### **Stratum 1**

Stratum 1 consists of the five largest Metropolitan Areas; the sampling objective was to select with certainty the five Metropolitan Areas in Stratum 1.

For Stratum 1, the probability of selection is 1 for each of the 5 Metropolitan Areas in the stratum. Therefore, the first-stage weight is also 1 for each MA in Stratum 1.

#### **Stratum 2**

Stratum 2 is comprised of the 8 remaining Metropolitan Areas in the target population; the sampling objective was to select with probability proportional to size (i.e., PPS) 4 Metropolitan Areas from Stratum 2.

For the PPS sample of 4 out of 8 Metropolitan Areas in stratum 2, the probability of selection of the  $j^{\text{th}}$  Metropolitan Area in stratum '2', say  $p_{2j}$ , may be represented as,

$$1) \quad p_{2j} = \frac{y_2 * m_{2j}}{M_2}$$

where,

$y_g$  - the number of sampled Metropolitan Areas in stratum  $g$  ( $g=2$  for Colombia) ,

$m_{gj}$  - the measure of size for the  $j^{\text{th}}$  PSU in stratum 2; in the case of the Colombia STEP Survey the measure of size for the first stage selection of PSUs in stratum 2 was the estimated number of persons aged 15-64 in each PSU in stratum 2 ,

$M_2$  -the estimated total number of persons aged 15-64 in stratum 2, based on the sample frame information, is represented by

$$2) \quad M_2 = \sum_1^{y_2} m_{2j}$$

Table 12 provides a summary of the sample design information for the first stage of sampling.

Table 12: First-stage Sample Design summary				
Stratum	Metropolitan Area (MA)	# of Sampled PSUs $y_g$	Population 15 -64 (estimated)	MA Selected
$g$			$m_{gj}$	
1 (Certainty inclusion)	Bogotá	5	4,387,012	Yes
	A.M Medellín		2,044,986	Yes
	A.M Cali		1,333,425	Yes
	A.M Barranquilla		959,843	Yes
	A.M Bucaramanga		605,385	Yes
Total – Stratum 1 ( $M_1$ )			9,330,651	
2 (Probabilistic - PPS)	Cartagena	4	544,336	No
	A.M Pereira		347,238	No
	A.M. Manizales		249,308	Yes
	Pasto		205,231	No
	A.M Cúcuta		422,460	Yes
	Villavicencio		215,640	Yes
	Montería		180,092	No
	Ibagué		285,455	Yes
Total Stratum 2 ( $M_2$ )			2,449,760	
Total (13 Areas)			11,780,411	

The basic first stage weight,  $W1_{gj}$ , for the  $j^{\text{th}}$  PSU in stratum 'g' is calculated as the inverse of its probability of selection:

$$3) \quad W1_{gj} = \begin{cases} 1, & \text{if } g = 1 \\ \frac{M_g}{y_g * m_{gj}}, & \text{if } g = 2 \end{cases}$$

➤ The basic first stage PSU weight is appended to the STEP data file.

**Example 1: First Stage Weight, 'Probabilistic-PPS' Stratum**

The basic first stage weight for Colombia MA A.M. Manizales, a PPS selection in stratum 2, is calculated as follows:

$$W1_{21} = \frac{1}{p_{21}} = \frac{M_2}{y_1 * m_{11}} = \frac{2,449,760}{4 * 249,308} \approx 2.46$$

The basic first stage weight is similarly calculated for each of the Metropolitan Areas in the 'Probabilistic-PPS' Stratum.

## 5.2 Second-stage Weight (SECTION Weight)

The calculation of the second stage weight is carried out for the selected Metropolitan Areas.

For the PPS sample of Sections, the probability of selection of the  $s^{\text{th}}$  Section in the  $j^{\text{th}}$  Metropolitan Area, say  $p_{gjs}$ , may be represented as,

$$4) \quad p_{gjs} = \begin{cases} \frac{y_{gjs} * m_{gjs}}{M_{gj}}, & \text{if } (y_{gjs} * m_{gjs}) < M_{gj} \\ 1, & \text{if } (y_{gjs} * m_{gjs}) \geq M_{gj} \end{cases}$$

where,

$y_{gj}$  - the number of sampled Sections in the  $j^{\text{th}}$  Metropolitan Area (MA),

$m_{gjs}$  - the measure of size for the  $s^{\text{th}}$  Section in the  $j^{\text{th}}$ ; in the case of the Colombia STEP Survey the measure of size for the second stage selection of Sections was the estimated number of persons aged 15-64 in each in each selected MA,

$M_{gj}$  - the total number of persons aged 15-64 in the  $j^{\text{th}}$  MA, based on the sample frame information

N.B. Based on the Colombia final data there are no Sections that were selected with certainty since  $(y_{gjs} * m_{gjs}) < M_{gj}$  for all sampled Sections .

The basic second stage weight,  $W2_{gjs}$ , for the  $s^{\text{th}}$  Section in the  $j^{\text{th}}$  Metropolitan Area is calculated as the inverse of its probability of selection:

$$5) \quad W2_{gjs} = \frac{1}{p_{gjs}} = \frac{M_{gj}}{y_{gjs} * m_{gjs}}$$

### **Example 2: Second Stage Basic Weight**

The basic second stage weight for Section #201-2 in the Colombia PSU # 2- Medellín, for which 44 Sections were chosen from 1326 Sections, is calculated as follows:

$$W2_{gjs} = \frac{m_{gj}}{n_{gj}} = \frac{2,044,986}{44 \times 2021} \approx 22.997$$

The basic second stage weight is similarly calculated for each of the 44 selected Sections.

## 5.2.1 Second-stage Response Adjustment

For treatment in the weighting process the data collection outcomes for the sampled Sections may be summarized in four possible categories: participant, ineligible, non-participant with known eligibility, non-participant with unknown eligibility.

- I. **Participant Section** - A participating Section 's' is an eligible sampled Section in which there are one or more Blocks that yields sufficient data to be considered a participant in the survey. The sampled Sections 'k' with Section-level result code 1 satisfies the requirement for classification of a selected Section as a survey participant. The Section result codes (variable 'SectBlk\_Status' in the Colombia data file) and corresponding code descriptions for participating Sections 'k' are as follows:

I. Section Result Codes (i.e., data file variable 'SectBlk_Status') – Participant, Eligible Section	
Code	Code Description
1	Complete Survey
2	Block out of universe
3	Block not activated
4	Block Refusal

- II. **Ineligible Section** – A sampled Section 'x' is categorized as ineligible if it was deemed to be outside the survey target population. The result code for ineligible sampled Sections 'x' is as follows:

II. Section Result Codes (i.e., data file variable 'SectBlk_Status') – Ineligible Section	
Code	Code Description
6	Section out of universe

- III. **Non-participant Section - Eligibility Known:** A non-participating Section 'z' results when an eligible Section does not participate due to refusal. The result code for sampled Sections 'z' with known eligibility is as follows:

III. Section Result Codes (i.e., data file variable 'SectBlk_Status') – Non-participant - Eligibility Known	
Code	Code Description
5	Section Refusal

- IV. **Non-participant Section - Eligibility Unknown:** A non-participating Section 'u' resulted when a selected Section was not activated (this pertains to the reserve sample cases). The result code for non-participating Sections 'u' with unknown eligibility is as follows:

IV. Section Result Codes (i.e., data file variable 'SectBlk_Status') – Non-participant - Eligibility Unknown	
Code	Code Description
7	Section Not Activated

The total number of sampled Sections in the  $j^{\text{th}}$  MA may be represented by the following equation.

$$6) \quad n_{js} = n_{jk} + n_{jx} + n_{jz} + n_{ju}$$

where,

- $n_{js}$  - the number of Sections sampled in MA 'j',
- $n_{jk}$  - the number of Sections 'k' that participate in MA 'j',
- $n_{jx}$  - the number of ineligible Sections 'x' in MA 'j',
- $n_{jz}$  - the number of non-participating Sections 'z' with known eligibility in MA 'j',
- $n_{ju}$  - the number of non-participating Sections 'u' with unknown eligibility in MA 'j'.

Table 13 provides the STEP counts from the Colombia data file for the Section outcome categories described above.

**Table 13: Colombia Sample Count of SECTIONS by Section Final Response Category by Metropolitan Area**

Metro Area	SECTION-Final Response Category ('RespCat_SECT')				
	1-Participant	2-Ineligible	3-Non-participant; Eligible	4-Non-activated; Unknown Eligibility	Total
(j)	( $n_{jk}$ )	( $n_{jx}$ )	( $n_{jz}$ )	( $n_{ju}$ )	( $n_{js}$ )
1-Bogota	83	0	8	1	92
2-Medellin A.M.	40	2	2	0	44
3-Cali A.M.	24	0	6	0	30
4-Barranquilla A.M.	17	0	5	1	23
5-Bucaramanga A.M.	13	0	2	1	16
7-Manizales A.M.	11	0	0	2	13
9-Villavicencio	10	0	1	1	12
11-Cucuta.	19	0	0	2	21
13-Ibague	14	0	0	2	16
Total	231	2	24	10	267



### 5.2.2 Adjustment of SECTION Weights for Unknown Eligibility

The STEP assumes that the SECTIONS 'u' with unknown eligibility in the  $j^{\text{th}}$  MA in stratum 'g' are comprised of some proportion, say  $\varepsilon_{js}$ , that are eligible SECTIONS and the complementary proportion,  $1 - \varepsilon_{js}$ , that are ineligible SECTIONS. Note that the calculation of  $\varepsilon_{js}$  is carried out independently for the SECTIONS in each selected metropolitan area.

Amongst the  $n_{ju}$  non-participating SECTIONS 'u' with unknown eligibility the proportion of SECTIONS with known eligibility is estimated to be

$$7) \quad \varepsilon_{js} = \frac{\text{Total \# of known eligible cases}}{\text{Total \# of known eligible \& known ineligible cases}} = \frac{n_{jk} + n_{jz}}{n_{jk} + n_{jx} + n_{jz}}$$

➤ Note that the eligible proportion is the same for all eligible participating SECTIONS in MA 'j'.

#### **Example 3: Eligible Proportion**

The counts in Table 13 provide input for calculation of the proportion of Sections with known eligibility.

For the Sections in Medellín, the calculation of the eligibility proportion is as follows,

$$\begin{aligned} \varepsilon_{12} &= \frac{\text{Total \# of known eligible cases}}{\text{Total \# of known eligible \& known ineligible cases}} \\ &= \frac{n_{1k} + n_{1z}}{n_{1k} + n_{1x} + n_{1z}} \\ &= \frac{40 + 2}{40 + 2 + 2} = \frac{42}{44} \approx 0.954545 \end{aligned}$$

### 5.2.3 Adjustment of SECTION Weights for Non-participation

A Section-level response adjustment is necessary to compensate for SECTIONS that did not participate in the survey.

The total number of sampled SECTIONS in MA 'j' is represented by the following equation.

$$8) \quad n_{js} = n_{jk} + n_{jx} + n_{jz} + n_{ju}$$

where,

- $n_{js}$  - the number of Sections sampled in MA 'j',
- $n_{jk}$  - the number of Sections 'k' that participate in MA 'j',
- $n_{jx}$  - the number of ineligible Sections 'x' in MA 'j',
- $n_{jz}$  - the number of non-participating Sections 'z' with known eligibility in MA 'j',
- $n_{ju}$  - the number of non-participating Sections 'u' with unknown eligibility in MA 'j'.

The SECTION-level response adjustment for the  $j^{\text{th}}$  MA,  $R_{js}$ , is calculated as follows:

$$9) \quad R_{js} = \frac{\text{Total eligible SECTIONS in MA 'j'}}{\text{Total participating SECTIONS in MA 'j'}} = \frac{n_{jk} + n_{jz} + (\varepsilon_j * n_{ju})}{n_{jk}}$$

- The response adjustment ratio,  $R_{js}$ , incorporates the adjustment for unknown eligibility,  $\varepsilon_{js}$ .
- The SECTIONS that are categorized as ineligible,  $n_{jx}$ , are excluded from the calculation of the SECTION-level response adjustment.

**Example 4: Second Stage Response Adjustment for sampled SECTIONS**

For example, the following results were obtained for Medellín

$n_{js}$	$n_{jk}$	$n_{jx}$	$n_{jz}$	$n_{ju}$
44	42	2	2	0

Therefore,

$$\begin{aligned} R_{12} &= \frac{\text{Total eligible households in PSU } j}{\text{Total participating households in PSU } j} = \frac{n_{jk} + n_{jz} + (\varepsilon_{12} * n_{ju})}{n_{jk}} \\ &= \frac{42 + 2 + (0.954545 * 0)}{42} \\ &\approx 1.05 \end{aligned}$$

The response adjustment ratio,  $R_{js}$ , is similarly calculated for each activated sampled SECTION.

## 5.2.4 Final Second-stage SECTION Weight

The final SECTION weight is an interim weight used in the calculation of the subsequent stage weights.

The final SECTION weight is calculated as the product of the Metropolitan Area weight, the basic Section weight, and the response adjustment for the basic Section weight. Therefore, for Section 's' in MA 'j', the final household weight,  $W2_s$ , is calculated as follows:

$$10) \quad W2_s = W1_{gj} * W2_{gjs} * R_{js}$$

- The final second stage SECTION weight is appended to the STEP data file.

**Example 5: Second Stage – Final SECTION Weight**

For the metropolitan area Medellín,

$$\begin{aligned} W2_s &= W1_j * W2_{gjs} * R_{js} \\ &= 1 * 22.997 * 1.05 \\ &\approx 24.147 \end{aligned}$$

The final Section weight is similarly calculated for all sampled Sections.

### 5.3 Third Stage: Block Weight

The calculation of the third stage weight is carried out for the sampled Blocks within the participant Sections; the third stage Block weight is zero for non-participant Sections.

For the PPS sample of Blocks, the probability of selection of the  $b^{\text{th}}$  Block in the  $s^{\text{th}}$  Section, say  $p_{gjsb}$ , may be represented as,

$$11) \quad p_{gjsb} = \begin{cases} \frac{y_{gjsb} * m_{gjsb}}{M_{gjs}}, & \text{if } (y_{gjsb} * m_{gjsb}) < M_{gjs} \\ 1, & \text{if } (y_{gjsb} * m_{gjsb}) \geq M_{gjs} \end{cases}$$

where,

$y_{gjsb}$  - the number of sampled Blocks in the  $s^{\text{th}}$  Section in the  $j^{\text{th}}$  Metropolitan Area (MA),

$m_{gjsb}$  - the measure of size for the  $b^{\text{th}}$  Block in the  $s^{\text{th}}$  Section in the  $j^{\text{th}}$  MA, in the case of the Colombia STEP Survey the measure of size for the third stage selection of Blocks was the estimated number of persons aged 15-64 in each in each selected Section,

$M_{gjs}$  - the total number of persons aged 15-64 in the  $s^{\text{th}}$  Section in the  $j^{\text{th}}$  MA, based on the sample frame information

N.B. Based on the Colombia final data there are some participant Sections that included one or more sampled Blocks that were selected with certainty since  $\frac{y_{gjsb} * m_{gjsb}}{M_{gjs}} \geq 1$ .

The basic third stage weight,  $W3_{gjsb}$ , for the  $b^{\text{th}}$  Block in the  $s^{\text{th}}$  Section in the  $j^{\text{th}}$  Metropolitan Area is calculated as the inverse of its probability of selection:

$$12) \quad W3_{gjsb} = \frac{1}{p_{gjsb}} = \begin{cases} \frac{M_{gjs}}{y_{gjsb} * m_{gjsb}}, & \dots \text{if } (y_{gjsb} * m_{gjsb}) < M_{gjs} \\ 1, & \dots \text{if } (y_{gjsb} * m_{gjsb}) \geq M_{gjs} \end{cases}$$

#### **Example 6:** Third Stage Basic Weight

The basic third stage weight for Block 3 in Section #201-2 in the Colombia Metro Area Medellín, for which 44 Sections were chosen from 1326 Sections, is calculated as follows:

$$W3_{gjsb} = \frac{M_{gjs}}{y_{gjsb} * m_{gjsb}} = \frac{2021}{4 \times 207} \approx 2.441$$

The basic third stage weight is similarly calculated for each of the selected Blocks.

### 5.3.1 Third Stage Response Adjustment

For the sampled Blocks from a participant Section, the data collection outcomes for the sampled Blocks may be summarized in four possible categories: participant, ineligible, non-participant with known eligibility, non-participant with unknown eligibility.

- I. **Participant Block** - A participating Block 'k' is an eligible sampled Block in which at least one complete survey was obtained. The result code (variable 'SectBlk\_Status' in the Colombia data file) and corresponding code description for participating Blocks 'k' is as follows:

I. Block Result Codes (i.e., data file variable 'SectBlk_Status') – Participant, Eligible Block	
Code	Code Description
1	Complete Survey

- II. **Ineligible Block** – A sampled Block 'x' is categorized as ineligible if it was deemed to be outside the survey target population. The result code for ineligible sampled Blocks 'x' is as follows:

II. Block Result Codes (i.e., data file variable 'SectBlk_Status') – Ineligible Block	
Code	Code Description
2	Block out of universe

- III. **Non-participant Block - Eligibility Known:** A non-participating Block 'z' results when an eligible Block does not participate due to refusal. The result code for sampled Blocks 'z' with known eligibility is as follows:

III. Block Result Codes (i.e., data file variable 'SectBlk_Status') – Non-participant Block, Eligibility Known	
Code	Code Description
4	Block Refusal

- IV. **Non-participant Block - Eligibility Unknown:** A non-participating Block 'u' resulted when a selected Block was not activated (this pertains to the reserve sample cases). The result code for non-participating Blocks 'u' with unknown eligibility is as follows:

IV. Block Result Codes (i.e., data file variable 'SectBlk_Status') – Non-participant Block, Eligibility Unknown	
Code	Code Description
3	Block not activated

The total number of sampled Blocks in the  $s^{\text{th}}$  sampled Section in the  $j^{\text{th}}$  MA may be represented by the following equation.

$$13) \quad n_{jsb} = n_{jsk} + n_{jsx} + n_{jsz} + n_{jsu}$$

where,

- $n_{jsb}$  - the number of Blocks sampled in the  $s^{\text{th}}$  sampled Section,
- $n_{jsk}$  - the number of Blocks 'k' that participate in the  $s^{\text{th}}$  sampled Section,
- $n_{jsx}$  - the number of ineligible Blocks 'x' in the  $s^{\text{th}}$  sampled Section,
- $n_{jsz}$  - the number of non-participating Blocks 'z' with known eligibility in the  $s^{\text{th}}$  sampled Section,
- $n_{jsu}$  - the number of non-participating Sections 'u' with unknown eligibility in the  $s^{\text{th}}$  sampled Section.

Table 14 provides the STEP sample counts from the Colombia data file for the Block outcome categories described above; note that table 12 provides the sample count of Blocks for participant Sections only.

**Table 14: Colombia Number of Sampled BLOCKs in Participant Sections by Block Final Response Category by Metropolitan Area**

Metro Area	Block-Final Response Category ('RespCat_BLK')				
	1-Participant	2-Ineligible	3-Non-participant; Eligible	4-Non-activated; Unknown Eligibility	Total
(j)	( $n_{jsk}$ )	( $n_{jsx}$ )	( $n_{jsz}$ )	( $n_{jsu}$ )	( $n_{js}$ )
1-Bogota	236	1	38	57	332
2-Medellin A.M.	90	4	3	63	160
3-Cali A.M.	54	0	6	36	96
4-Barranquilla A.M.	37	0	4	27	68
5-Bucaramanga A.M.	28	2	7	15	52
7-Manizales A.M.	21	1	1	21	44
9-Villavicencio	20	0	0	20	40
11-Cucuta.	46	2	1	27	76
13-Ibague	26	0	2	28	56
Total	558	10	62	294	924

### 5.3.2 Adjustment of Block Weights for Unknown Eligibility

The STEP assumes that the Blocks, and also Block Segments, 'u' with unknown eligibility in the  $s^{\text{th}}$  Section in the  $j^{\text{th}}$  MA in stratum 'g' are comprised of some proportion, say  $\epsilon_{jsb}$ , that are eligible SECTIONS and the complementary proportion,  $1 - \epsilon_{jsb}$ , that are ineligible SECTIONS. Note that the calculation of  $\epsilon_{jsb}$  is carried out independently for the Blocks in each selected Section.

Amongst the  $n_{jsu}$  non-participating BLOCKs 'u' with unknown eligibility the proportion of Blocks with known eligibility is estimated to be

$$14) \quad \epsilon_{jsb} = \frac{\text{Total \# of known eligible cases}}{\text{Total \# of known eligible \& known ineligible cases}} = \frac{n_{jsk} + n_{jsz}}{n_{jsk} + n_{jsx} + n_{jsz}}$$

- Note that the eligible proportion is the same for all eligible participating Blocks in the  $s^{\text{th}}$  Section.
- $\epsilon_{jsb} \geq 0.5$  since there are 4 sampled Blocks in each selected Section.

The Blocks with Unknown Eligibility, i.e.  $n_{jsu}$ , were redistributed to the other response categories as follows:

$$15) \quad \begin{cases} \text{If } \epsilon_{jsb} > 0.5, & n'_{jsz} = n_{jsz} + n_{jsu} \\ \text{If } \epsilon_{jsb} = 0.5, & \begin{cases} n'_{jsz} = n_{jsz} + n_{jsu} * \epsilon_{jsb} \\ n'_{jsx} = n_{jsx} + n_{jsu} * \epsilon_{jsb} \end{cases} \end{cases}$$

### 5.3.3 Adjustment of BLOCK Weights for Non-participation

A BLOCK-level response adjustment within the selected Section is necessary to compensate for BLOCKs that did not participate in the survey.

The BLOCK-level response adjustment depends on whether or not a selected Section included one or more Blocks selected with certainty.

#### Case 1: Section has no Blocks selected with certainty

After the redistribution of the Blocks with Unknown Eligibility, i.e.  $n_{jsu}$ , to the other response categories, the total number of sampled BLOCKs in the  $s^{th}$  Section in MA 'j' is represented by the following equation.

$$16) \quad n_{jsb} = n_{jsk} + n'_{jsx} + n'_{jsz}$$

where,

- $n_{jsb}$  - the number of Blocks sampled in the  $s^{th}$  Section.
- $n_{jsk}$  - the number of Blocks 'k' that participate in the  $s^{th}$  Section .
- $n'_{jsx}$  - the revised (i.e., after redistribution of the  $n_{jsu}$  cases) number of ineligible Blocks 'x' in the  $s^{th}$  Section.
- $n'_{jsz}$  - the revised (i.e., after redistribution of the  $n_{jsu}$  cases) number of non-participating Blocks 'z' with known eligibility in the  $s^{th}$  Section.

The BLOCK-level response adjustment for the  $s^{th}$  Section in the  $j^{th}$  MA,  $R_{jsb}$ , is calculated as follows:

$$17) \quad R_{jsb} = \frac{\text{Total eligible BLOCKs in Section 's' in MA 'j'}}{\text{Total participating BLOCKs in Section 's' in MA 'j'}} = \frac{n_{jsk} + n'_{jsz}}{n_{jsk}}$$

#### **Notes:**

- The response adjustment ratio,  $R_{jsb}$ , incorporates the adjustment for unknown eligibility.
- The SECTIONS that are categorized as ineligible,  $n_{jsx}$ , are excluded from the calculation of the BLOCK-level response adjustment.
- Within a selected Section,

$$M_{gjs} = \sum_1^{n_{jsb}} (W3_{gjsb} * R_{jsb} * m_{gjsb})$$

In other words, the weighted sum of the persons aged 15 to 64 in the selected Blocks within a Section must equal the number of persons aged 15 to 64 within the selected section.

#### Case 2: Section has one or more Blocks selected with certainty

The sampled Blocks were grouped according to PPS selection Blocks and certainty selection Blocks in order to carry out the Block non-response adjustment within a selected Section. The response adjustment scenarios are as follows:

1) The Block non-response involves only PPS Blocks.

In this case, the non-response adjustment for the PPS Blocks in the Section is

$$18) \quad R_{jsb} = \frac{\text{Total eligible BLOCKs in Section 's' in MA 'j'}}{\text{Total participating BLOCKs in Section 's' in MA 'j'}} = \frac{n_{jsk} + n'_{jsz}}{n_{jsk}}$$

where  $n_{jsk}$  ,  $n'_{jsz}$  , and  $n_{jsk}$  pertain to only the PPS Blocks within the Section.

2)The Block non-response involves only Blocks selected with certainty and none of these Blocks are participants.

In this case, the PPS Blocks are adjusted so that they also represent the non-respondent certainty Blocks; the non-response adjustment for the PPS Blocks in the Section is

$$19) \quad R_{jsb} = \frac{\# \text{ of persons aged 15 to 64 in Section}}{\# \text{ of persons aged 15 to 64 in PPS Blocks}} = \frac{M_{gjs}}{\sum_1^{n_{jsk}} W3_{gjsb} * m_{gjsb}}$$

where  $n_{jsk}$  ,  $W3_{gjsb}$  , and  $m_{gjsb}$  pertain to only the PPS Blocks within the Section, and  $M_{gjs}$  is the number of persons aged 15-64 in the Section.

3)The Block non-response involves only Blocks selected with certainty and at least one of these Blocks is a participant (Note, in this case, there is more than one certainty Block selected in the Section.)

a. The non-response adjustment for the certainty Blocks in the Section is

$$20) \quad R_{jsb} = \frac{\# \text{ of persons 15 to 64 in Section}}{\# \text{ of persons 15 to 64 in Participant certainty Blocks}} = \frac{M_{gjs}}{\sum_1^{n_{jsk}} W3_{gjsb} * m_{gjsb}}$$

where  $n_{jsk}$  ,  $W3_{gjsb}$  , and  $m_{gjsb}$  pertain to only the certainty Blocks within the Section, and  $M_{gjs}$  is the number of persons aged 15-64 in the Section.

b. The above non-response adjustment is also applicable when all Blocks within a Section were selected with certainty.

4)The Block non-response involves both PPS Blocks and Blocks selected with certainty, i.e., there is at least one non-responding Block amongst the PPS Blocks and there is also at least one non-responding Block amongst the certainty Blocks (Note, in this case, there is more than one certainty Block selected in the Section.)

a. The non-response adjustment for the PPS Blocks in the Section is

$$24) \quad R_{jsb} = \frac{\text{Total eligible BLOCKS in Section 's' in MA 'j'}}{\text{Total participating BLOCKS in Section 's' in MA 'j'}} = \frac{n_{jsk} + n'_{jsz}}{n_{jsk}}$$

where  $n_{jsk}$  ,  $n'_{jsz}$  , and  $n_{jsk}$  pertain to only the PPS Blocks within the Section.

a. The non-response adjustment for the certainty Blocks in the Section is

$$25) \quad R_{jsb} = \frac{\# \text{ of persons 15 to 64 in Section}}{\# \text{ of persons 15 to 64 in Participant certainty Blocks}} = \frac{M_{gjs}}{\sum_1^{n_{jsk}} W3_{gjsb} * m_{gjsb}}$$

where  $n_{jsk}$  ,  $W3_{gjsb}$  , and  $m_{gjsb}$  pertain to only the certainty Blocks within the Section, and  $M_{gjs}$  is the number of persons aged 15-64 in the Section.

### 5.3.4 Third Stage Response-Adjusted BLOCK Weight

The response-adjusted BLOCK weight is an interim weight used in the calculation of the subsequent fourth stage weight.

The response-adjusted BLOCK weight is calculated as the product of the final Section Weight, the basic Block weight adjusted for the multiple Block Segment factor, and the response adjustment for the basic Block weight. Therefore, for Block 'b' in Section 's' in MA 'j', the response adjusted Block weight,  $W3_{Radj}$ , is calculated as follows:

$$26) \quad W3_{Radj} = W3_{gjsb} * R_{jsb}$$

➤ The final third stage BLOCK weight is appended to the STEP data file.

#### **Example 7:** Third Stage – Response-adjusted Block Weight

For Block #201-2-3 in Medellín,

$$\begin{aligned} W3_{Radj} &= W3_{gjsb} * R_{jsb} \\ &= 2.441 \times 2.0 \\ &\approx 4.882 \end{aligned}$$

The final Block weight is similarly calculated for all sampled Blocks.

### 5.4 Fourth Stage: Block Segment Weight (Adjustment for Multiple Block Segments)

Some selected Blocks contained 80 or more dwellings. In such cases, the Blocks were divided into Block Segments and one Block Segment was randomly selected. Therefore, for each Block 'b', the basic Block weight is adjusted by the following multiple Block Segment factor,

$$24) \quad BS_{gjsb} = \frac{b'_{gjsb}}{n_{gjsb}}$$

where,

$b'_{gjsb}$  - the number of Block Segments in Block 'b',

$n_{gjsb}$  - the number of Block Segments sampled in Block 'b', (i.e.,  $n_{gjsb} = 1$  for all sampled Blocks).

Therefore the Block Weight, adjusted for the number of Block segments, is represented by

$$25) \quad W4_{gjsb} = W3_{Radj} * BS_{gjsb}$$



**Example 8: Basic Block Weight Adjustment for Multiple Block Segments**

The basic third/fourth stage weight for sampled Block 3 in Sector-Section 201-2 in Medellín, which contains 3 Block segments, is calculated as follows:

$$W4_{gjsb} = W3_{Radj} * \frac{b'_{gjsb}}{n_{gjsb}} = 4.882 * \frac{3}{1} \approx 14.645$$

#### 5.4.1 Final Third/Fourth Stage BLOCK-Segment Weight

The final BLOCK-Segment weight is an interim weight used in the calculation of the subsequent stage weights.

The final BLOCK-Segment weight is calculated as the product of the final Section Weight, the basic Block weight adjusted for the multiple Block Segment factor and the response adjustment for the basic Block weight. Therefore, for Block 'b' in Section 's' in MA 'j', the final Block weight,  $W_B$ , is calculated as follows:

$$26) \quad W4_B = W2_S * W4_{gjsb}$$

- The final third/fourth stage BLOCK weight is appended to the STEP data file.

**Example 9: Third/Fourth Stage – Final Block-Segment Weight**

For Block #201-2-3 in Medellín,

$$\begin{aligned} W4_B &= W2_S * W4_{gjsb} \\ &= 24.147 \times 14.645 \\ &\approx 353.629 \end{aligned}$$

The final Block-Segment weight is similarly calculated for all sampled Blocks/Segments.

## 5.5 Dwelling Weight – Fifth Stage

### 5.5.1 Dwelling Weight - Basic

The calculation of the fifth stage weight is carried out for sampled Dwellings from the participant Blocks; the fifth stage Dwelling weight is zero for non-participant Blocks.

The basic weight for the  $d^{\text{th}}$  dwelling in Block 'b' in Section 's' in MA 'j',  $W5_{gjsbd}$ , is calculated as follows:

$$27) \quad W5_{gjsbd} = \frac{m'_{gjsb}}{m_{gjsb}}$$

where,

$m'_{gjsb}$  - the number of **listed** dwellings in Block 'b' in Section 's' in MA 'j',

$m_{gjsb}$  - the number of dwellings sampled in Block 'b' in Section 's' in MA 'j'.

#### **Example 10:** Fifth Stage Basic Weight

The basic third stage weight for each of the 9 sampled dwellings in Block #201-2-3 in Medellín, which contains 36 **listed** households, is calculated as follows:

$$W5_{gjsbd} = \frac{m'_{1js}}{n_{1js}} = \frac{36}{9} = 4.0$$

The basic fifth stage weight is similarly calculated for each of the sampled dwellings.

### 5.5.2 Dwelling Weight – Adjustment for Observed Eligibility

Prior to any contact with a sampled dwelling the eligibility of the dwelling may be determined based on the dwelling's physical status, e.g., the dwelling exists or the dwelling no longer exists or the dwelling is no longer habitable. Therefore, an adjustment for non-response at the dwelling level may be carried out based on the observed condition of the dwelling.

Thus, for the sampled Dwellings from a participant Block, the data collection outcomes for the sampled dwellings may be summarized in two possible categories: eligible, ineligible.

- I. **Eligible Dwelling** - An eligible dwelling 'k' is any sampled dwelling that exists, is habitable, and/or is deemed to be occupied; in other words, an eligible dwelling is a sampled dwelling that is not categorized as ineligible based on the physical condition of the dwelling. The sampled dwellings 'k' with result of interview codes other than 07-09 satisfy the requirement for classification of the dwelling as an eligible dwelling.

- II. **Ineligible** – A sampled dwelling ‘x’ is categorized as ineligible if it no longer exists, is uninhabitable, or is vacant. The result codes for ineligible sampled dwellings ‘x’ are as follows:

#### I. Dwelling-level Codes – Ineligible Dwelling

Code	Code Description
07	Dwelling could not be found / selected address has no household
08	Vacant dwelling
09	Dwelling not habitable/dwelling destroyed/dwelling converted to commercial use

A dwelling-level response adjustment is necessary to compensate for dwellings that were not eligible to participate in the survey. The dwelling-level response adjustment for the  $d^{\text{th}}$  dwelling in the  $b^{\text{th}}$  Block in Section ‘s’ in MA ‘j’,  $R_{jsbd}$ , is calculated as follows:

$$28) \quad R_{jsbd} = \begin{cases} 1, & \text{Eligible Dwelling} \\ 0, & \text{Ineligible Dwelling} \end{cases}$$

- Essentially, the dwellings that are categorized as ineligible are excluded from the calculation of the dwelling-level response adjustment.

### 5.5.3 Dwelling Weight - Final

The final dwelling weight is appended to the STEP data file; it is an essential component of the benchmark adjustment (i.e., adjustment to known population totals) for the household weights and the person weights.

The final dwelling weight is calculated as the product of the final Block weight, the basic dwelling weight, and the response adjustment for the basic dwelling weight. Therefore, for the  $d^{\text{th}}$  dwelling in the  $b^{\text{th}}$  participant Block in Section ‘s’ in MA ‘j’, the final dwelling weight,  $W_D$ , is calculated as follows:

$$29) \quad W5_D = W4_B * W5_{gjsbd} * R_{jsbd}$$

- The final fifth stage household weight is appended to the STEP data file.

#### **Example 11:** Fifth Stage – Final Dwelling Weight

For Block #201-2-3 in Medellín,

$$\begin{aligned} W5_D &= W4_B * W5_{gjsbd} * R_{jsbd} \\ &= 353.629 \times 4 \times 2.83 \times 1 \\ &\approx 1414.516 \end{aligned}$$

The final dwelling weight is similarly calculated for all sampled dwellings.

## 5.6 Household Weight – Sixth Stage

### 5.6.1 Household Weight - Basic

The calculation of the sixth stage household weight is carried out for the sampled Households within the Eligible Dwellings from the fifth sampling stage; the sixth stage Household weight is zero for ineligible Dwellings.

Some selected dwellings were found to include more than one household. In such cases, one household was randomly selected for an interview. Therefore, for each eligible dwelling, 'd', the basic household weight is,

$$30) \quad W6_{dh} = \frac{h_d}{n_{dh}}$$

where,

$h_d$  - the number of actual households in dwelling 'd',

$n_{dh}$  - the number of households sampled in dwelling 'd', (i.e.,  $n_{dh} = 1$  for all sampled cases).

## 5.6.2 Household Weight – Adjustment for Non-Response

For the sampled Households within an eligible Dwelling, the data collection outcomes for the sampled households may be summarized in four possible categories: participant, ineligible, non-participant with known eligibility, non-participant with unknown eligibility.

- I. **Participant** - A participating household 'k' is an eligible sampled household that provides sufficient data to be considered a participant in the survey. The sampled households 'k' with person-level result codes 32-40, 51-78 and 98-99 satisfy the requirement for classification of the household as a survey participant. The result codes and corresponding code descriptions for participating households 'k' are as follows:

I. Household-level Codes – Participant, Eligible Household	
Code	Code Description
<b>Individual modules 2-7 <u>not</u> begun</b>	
32	Refusal by Selected Person to begin modules 2-7
33	Unusual circumstance - Selected Person (e.g., death in family, illness, fire in dwelling, etc.)
34	Refusal - Another household member refused to let selected individual do individual modules
35	Unable to contact Selected Person after three visits during field period
36	Temporarily absence/unavailability of Selected Person for field period (info from household member)
37	Hearing Problem - Selected Person is deaf or so hard-of-hearing, no translator available
38	Learning or mental disability (includes emotional conditions like severe depression) of Selected Person
39	Speech impairment of Selected Person prevented participation, no translator available
40	Language problem – Selected Person's language not understood by survey team, no translator available
<b>Individual modules 2-7 begun but not completed</b>	
51	Refusal by Selected Person to continue modules 2-7
52	Unusual circumstance - Selected Person
<b>General Booklet (module 9) not begun</b>	
61	Unable to read or write in language of General Booklet – Selected Person refused to begin
62	Refusal by Selected Person to begin General Booklet
63	Unusual circumstance - Selected Person
64	Blindness or visual impairment of Selected Person prevented completion of General Booklet
65	Physical disability of Selected Person prevented completion of General Booklet
<b>General booklet (module 9) looked through or attempted but not completed</b>	
71	Selected Person looked at Booklet but unable to read & write the language of Booklet - did not begin
72	Refusal by Selected Person to continue the General Booklet
73	Interruption – unacceptably long
74	Unusual circumstance - Selected Person
75	Attempted General Booklet but no answers marked
76	Marked some items but did not attempt all items in General Booklet
77	Marked some items & attempted all items in General Booklet
78	Marked all items in General Booklet
98	Sample selection error – an incorrect eligible person was selected by the interviewer
99	Sample selection error – an ineligible person was selected by the interviewer

- II. **Ineligible** – A sampled household 'x' is categorized as ineligible if there is no household member in the eligible 15-64 age range. The result codes for ineligible sampled households 'x' are as follows:

II. Household-level Codes – Ineligible Household	
Code	Code Description
31	No household member in the eligible range of 15-64

III. **Non-participant - Eligibility Known:** A non-participating household ‘z’ results when an eligible household does not participate due to refusal, unusual circumstance, or the household was selected in error. The result codes for sampled households ‘z’ with known eligibility are as follows:

III. Household-level Codes – Non-participants, Eligible Household	
Code	Code Description
21	Household Refusal to continue Household Module (i.e., begun but not complete)
22	Household unusual circumstance (e.g., death in family, illness, fire in dwelling, etc.) prevented continuance
94	Case dropped due to data quality issue

IV. **Non-participant - Eligibility Unknown:** A non-participating household ‘u’ results when a household does not participate due to non-contact with a responsible household member, inconclusive information about the eligibility status of the sampled household, or sample case was not activated (this usually pertains to the reserve sample cases). The result codes for non-participating households ‘u’ with unknown eligibility are as follows:

IV. Household-level Codes – Non-participants, Unknown Eligibility Household	
Code	Code Description
01	Household Refusal
02	Household unusual circumstance (e.g., death in family, illness, fire in dwelling, etc.)
03	No knowledgeable household member found after 3 visits (only a child, non-competent adult, etc.)
04	Temporarily absent/unavailable for field period (information from other persons)
05	No competent household member to interview (because of severe illness, mental disability, etc.)
06	Language problem – no household member spoke a language understandable by survey team, no translator available
10	No Answer at door
95	Sample case not activated
96	General non-response; Reason unspecified by survey firm
97	Sample selection error – a wrong household was selected in the field

The non-response adjustments to the basic household weight were carried out at the Block level. The total number of sampled households in each sampled Block may be represented by the following equation.

$$31) \quad n_{bdh} = n_{bdk} + n_{bdx} + n_{bdz} + n_{bdu}$$

where,

- $n_{bdh}$  - the number of households sampled in Block ‘b’,
- $n_{bdk}$  - the number of households ‘k’ that participate in Block ‘b’,
- $n_{bdx}$  - the number of ineligible households ‘x’ in Block ‘b’,
- $n_{bdz}$  - the number of non-participating households ‘z’ with known eligibility in Block ‘b’,
- $n_{bdu}$  - the number of non-participating households ‘u’ with unknown eligibility in Block ‘b’.

### 5.6.3 Household Weight – Adjustment for Unknown Eligibility

The STEP assumes that the households ‘u’ with unknown eligibility in the  $b^{\text{th}}$  Block are comprised of some proportion, say  $\epsilon_{bdh}$ , that are eligible households and the complementary proportion,  $1 - \epsilon_{bdh}$ , that are ineligible households.

Amongst the  $n_{bdu}$  non-participating households ‘u’ with unknown eligibility the proportion of households with known eligibility is estimated to be

$$32) \quad \epsilon_{bdh} = \frac{\text{Total \# of known eligible cases in sampled Block}}{\text{Total \# of known eligible \& known ineligible cases in sampled Block}} = \frac{n_{bdk} + n_{bdz}}{n_{bdk} + n_{bdx} + n_{bdz}}$$

- Note that the eligible proportion is the same for all eligible participating households in a sampled Block.

#### **Example 12:** Eligible Proportion of Households in a Block

For Block #201-2-3 in Medellín,

$n_{bdh}$	$n_{bdk}$	$n_{bdx}$	$n_{bdz}$	$n_{bdu}$
17	15	0	0	2

Therefore, the calculation of the eligibility proportion is as follows,

$$\begin{aligned} \epsilon_{bdh} &= \frac{\text{Total \# of known eligible cases}}{\text{Total \# of known eligible \& known ineligible cases}} \\ &= \frac{n_{jsk} + n_{jsz}}{n_{jsk} + n_{jsx} + n_{jsz}} = \frac{15 + 0}{15 + 0 + 0} = \frac{15}{15} = 1.0 \end{aligned}$$

The Households with Unknown Eligibility, i.e.  $n_{bdu}$ , were redistributed to the other response categories as follows:

$$33) \quad \begin{cases} \text{If } \epsilon_{bdh} > 0.5, & n'_{bdz} = n_{bdz} + n_{bdu} \\ \text{If } \epsilon_{bdh} = 0.5, & \begin{cases} n'_{bdz} = n_{bdz} + n_{bdu} * \epsilon_{bdh} \\ n'_{bdx} = n_{bdx} + n_{bdu} * \epsilon_{bdh} \end{cases} \end{cases}$$

#### **Example 13:** Re-distribution of Households with Unknown Eligibility in a Block

For Block #201-2-3 in Medellín, the Households with Unknown Eligibility were redistributed to  $n_{bdz}$  since  $\epsilon_{bdh} = 1.0$  for this Block, i.e.,

$$n'_{bdz} = n_{bdz} + n_{bdu} = 0 + 2 = 2.0$$

### 5.6.4 Household Weight - Adjustment for Non-participation

A household-level response adjustment is necessary to compensate for households that did not participate in the survey. The household-level response adjustment for the  $b^{\text{th}}$  Block,  $R_{bdh}$ , is calculated as follows:

$$34) \quad R_{bdh} = \frac{\text{Total eligible households in Block } b}{\text{Total participating households in Block } b} = \frac{n_{bdk} + n'_{bdz}}{n_{bdk}}$$

- The response adjustment ratio,  $R_{bdh}$ , incorporates the adjustment for unknown eligibility,  $\varepsilon_{bdh}$ .
- The households that are categorized as ineligible,  $n'_{bdx}$ , are excluded from the calculation of the household-level response adjustment.

#### **Example 14:** Second Stage Response Adjustment for sampled Households

For example, after redistribution of the Households with Unknown Eligibility to the other response categories the following results were obtained for Block #201-2-3 in Medellín,

$n_{bdh}$	$n_{bdk}$	$n'_{bdx}$	$n'_{bdz}$
17	15	0	2

Therefore,

$$\begin{aligned} R_{bdh} &= \frac{\text{Total eligible households in Block 'b'}}{\text{Total participating households in Block 'b'}} = \frac{n_{bdk} + n'_{bdz}}{n_{bdk}} \\ &= \frac{15 + 2}{15} \\ &\approx 1.13 \end{aligned}$$

The response adjustment ratio,  $R_{bdh}$ , is similarly calculated for all sampled Blocks.

### 5.6.5 Household Weight - Final

The final household weight is the weight used for producing household-level estimates, such as the proportion of households with a specific attribute. This weight is appended to the STEP data file.

The final household weight is calculated as the product of the final Dwelling weight, the basic household weight, and the response adjustment for the basic household weight. Therefore, for household 'k' in PSU 'j', the final household weight,  $W_H$ , is calculated as follows:

$$35) \quad W6_H = W5_D * W6_{dh} * R_{bdh}$$

- The final sixth stage household weight is appended to the STEP data file.

#### **Example 15:** Fifth/Sixth Stage – Final Household Weight

For Block #201-2-3 in Medellín,

$$\begin{aligned} W6_H &= W5_D * W6_{dh} * R_{bdh} \\ &= 1414.516 \times 1 \times 1.13 \\ &\approx 1603.118 \end{aligned}$$

The final household weight is similarly calculated for all sampled households.



## 5.7 Seventh-stage Weight (Person Weight)

### 5.7.1 Person Weight - Basic

The calculation of the seventh stage person weight is carried out for the participant Households from the sixth sampling stage; the seventh stage Person weight is zero for non-participant Households.

The basic seventh stage weight for the  $i^{\text{th}}$  person,  $W7_{hi}$ , is calculated as follows:

$$36) \quad W7_{hi} = \frac{a_{hi}}{n_{hi}} = \frac{a_{hi}}{1} = a_{hi}$$

where,

- $a_{hi}$  - the number of eligible persons in the  $h^{\text{th}}$  household in Block 'b' in Section 's' in MA 'j',
- $n_{hi}$  - the number of persons sampled in the  $h^{\text{th}}$  household in Block 'b' in Section 's' in MA 'j', (i.e.,  $n_{bdhi} = 1$  person in each Colombia sampled household).

$$\text{N.B.} \quad A_{hi} = \sum_{h=1}^{n_{bdk}} a_{hi} = \text{total \# of selected persons in the } b^{\text{th}} \text{ Block}$$

where,

$n_{bdk}$  - the number of participating households 'k' in the  $b^{\text{th}}$  Block

#### **Example 16:** Third Stage – Basic Person Weight within sampled Household

For Block #201-2-3 in Medellín, selected household # 1 has 2 eligible persons. Therefore,

$$W7_{hi} = \frac{a_{hi}}{n_{hi}} = \frac{2}{1} = 2$$

## 5.7.2 Seventh-stage Person-level Response Adjustment

For eligible participating households, there are three possible person-level outcomes for a Selected Person (SP), (i) a selected eligible person participates in the survey, (ii) a selected eligible person does not participate, (iii) a selected person is ineligible.

- I. **Participant SP** - A participating selected person 'p' is an eligible sampled person that provides sufficient data to be considered a participant in the survey. The sampled persons 'p' with result codes 51-88 satisfy the requirement for classification of the person as a survey participant. The result codes for participating persons 'p' are as follows:

<b>I. Person-level Codes – Participant Selected Person</b>	
<b>Code</b>	<b>Code Description</b>
<b>Individual modules 2-7 begun but not completed</b>	
51	Refusal by Selected Person to continue modules 2-7
52	Unusual circumstance - Selected Person
<b>General Booklet (module 9) not begun</b>	
61	Unable to read or write in language of General Booklet – Selected Person refused to begin
62	Refusal by Selected Person to begin General Booklet
63	Unusual circumstance - Selected Person
64	Blindness or visual impairment of Selected Person prevented completion of General Booklet
65	Physical disability of Selected Person prevented completion of General Booklet
<b>General booklet (module 9) looked through or attempted but not completed</b>	
71	Selected Person looked at Booklet but unable to read & write the language of Booklet - did not begin
72	Refusal by Selected Person to continue the General Booklet
73	Interruption – unacceptably long
74	Unusual circumstance - Selected Person
75	Attempted General Booklet but no answers marked
76	Marked some items but did not attempt all items in General Booklet
77	Marked some items & attempted all items in General Booklet
78	Marked all items in General Booklet
81	No attempt at any of the exercise booklet items / refusal to start
82	Interruption too long during exercise booklet administration
83	Unusual circumstance, Exercise Booklet administration terminated
84	Attempt at some Exercise Booklet items but no answers marked
85	Attempt at all Exercise Booklet items but no answers marked
86	Marked some Exercise Booklet items, not all items attempted
87	Marked some Exercise Booklet items, all items attempted
88	Marked all Exercise Booklet items

- II. **Non-participant SP** - A non-participating person ‘q’ results when an eligible person does not participate due to refusal, unusual circumstance, non-contact, language problem, physical disability (hearing, vision, other) or the person was selected in error. The result codes for non-participating persons ‘q’ are as follows:

II. Person-level Codes – Non-participant Selected Person	
Code	Code Description
<b>Individual modules 2-7 <u>not</u> begun</b>	
32	Refusal by Selected Person to begin modules 2-7
33	Unusual circumstance - Selected Person (e.g., death in family, illness, fire in dwelling, etc.)
34	Refusal - Another household member refused to let selected individual do individual modules
35	Unable to contact Selected Person after three visits during field period
36	Temporarily absence/unavailability of Selected Person for field period (info from household member)
37	Hearing Problem - Selected Person is deaf or so hard-of-hearing, no translator available
38	Learning or mental disability (includes emotional conditions like severe depression) of Selected Person
39	Speech impairment of Selected Person prevented participation, no translator available
40	Language problem – Selected Person’s language not understood by survey team, no translator available
94	Case dropped due to data quality issue
98	Sample selection error – an incorrect eligible person was selected by the interviewer

- III. **Ineligible SP** - A sampled person ‘f’ is categorized as ineligible if the person is not a member of the target population, for example, the person’s age is outside the eligible 15-64 age range. Ineligible sampled persons ‘f’ are assigned the result code 99.

III. Person-level Codes – Ineligible Selected Person	
Code	Code Description
99	Sample selection error – an ineligible person was selected by the interviewer

The third stage person weight is only relevant for sampled households in the final response category RespCat\_HH=1, i.e., participating households. Table 14 provides the pertinent frequency counts of the Final Result Codes.

**Table 14: Final Result Code for Participating Households (i.e., Households with FinRespCat\_HH=1)**

Final Result Code & Description	Frequency	Percent
32-Refusal by Selected Person to begin modules 2-7	236	8.4
33-Interview termination-SP unusual circumstance	8	.3
34-Refusal-Other Household Member refused to let SP participate	34	1.2
35-Unable to contact SP after 3 attempts	86	3.0
36-SP absent for entire survey period	15	.5
38-SP learning or mental disability	4	.1
39-SP speech impairment	2	.1
51-Refusal by SP to continue modules 2-7	23	.8
52-Interview termination-SP unusual circumstance	4	.1
61- Unable to read or write in language of General Booklet – SP refused to begin	22	.8
62-SP refusal to begin GB	29	1.0
63- Unusual circumstance - Selected Person	1	.0
64-SP blindness or visual impairment-unable to do GB	3	.1
71-SP looked through GB-unable to read & write language of GB	8	.3
72-SP refused to continue GB after starting	53	1.9
75-Attempted General Booklet but no answers marked	7	.2
76-Attempted part or all of General Booklet; less than 3 correct Core answers	318	11.3
81-No attempt at any of the exercise booklet items / refusal to start	27	1.0
83-Unusual circumstance, Exercise Booklet administration terminated	3	.1
84-Attempt at some Exercise Booklet items but no answers marked	6	.2
85-Attempt at all Exercise Booklet items but no answers marked	6	.2
86-Marked some Exercise Booklet items, not all items attempted	56	2.0
87-Marked some Exercise Booklet items, all items attempted	847	30.0
88-Marked all Exercise Booklet items	1021	36.1
98-Sample selection error – incorrect eligible person selected	6	.2
<b>Total</b>	<b>2825</b>	<b>100.0</b>

In each participating household in the STEP Survey, one eligible person was selected to be interviewed. Therefore, for each participating household the basic person weight,  $W7_{hi}$ , would also be the final person weight within a sampled Block if all persons in the Block participated. However, when there is one or more non-participating selected persons within a sampled Block, the basic person-weights for the Block require a response adjustment so that the final person weight for each participating person in the Block is also representative of the non-participating person(s) in the sampled households in the Block. In other words, the total number of participating persons in a sampled Block must represent the total number of eligible persons in the sampled households in the Block.

Therefore, within each sampled Block, a person-level response adjustment to the basic person weight is necessary to compensate for persons that did not participate in the survey.

The total number of eligible persons in each selected Block may be represented by the following equation:

$$37) \quad A_{bh} = A_{bhp} + A_{bhq}$$

where,

$$\begin{aligned} A_{bh} &= \sum_{h=1}^{n_{bk}} a_{hi} & - & \text{the total number of eligible persons for all } n_{hi} \text{ participating households} \\ & & & \text{in the } b^{\text{th}} \text{ Block,} \\ A_{bhp} &= \sum_{h=1}^{n_{bk}} a_{hp} & - & \text{the total number of participating eligible persons 'p' in all } n_{bdk} \\ & & & \text{participating households in the } b^{\text{th}} \text{ Block,} \\ A_{bhq} &= \sum_{h=1}^{n_{bk}} a_{hq} & - & \text{the total number of non-participating eligible persons 'q' for all } n_{jsb} \\ & & & \text{participating households in the } b^{\text{th}} \text{ Block,} \end{aligned}$$

Therefore, the person-level response adjustment for the  $b^{\text{th}}$  Block,  $R_{bhi}$ , is calculated as follows:

$$38) \quad R_{bhi} = \frac{\text{Total eligible persons in all participating sampled households in Block } b}{\text{Total participating persons in all participating sampled households in Block } b} = \frac{A_{bh}}{A_{bhp}}$$

Note that ineligible persons are excluded from the above ratio adjustment for person-level non-response.

**Example 17: Third Stage – Basic Person Weight Response Adjustment**

- i) For Block #201-2-3 in Medellín, there are 18 eligible persons across all participating households. Also, all selected persons participated in the survey and thus the selected persons represent the 18 eligible persons in the participating households. Therefore,

$$A_{bh} = \sum_{h=1}^{n_{bk}} a_{hi} = 18 ; \quad A_{bhp} = \sum_{h=1}^{n_{bk}} a_{hp} = 18$$

$$R_{bhi} = \frac{A_{bh}}{A_{bhp}} = \frac{18}{18} = 1.00$$

- ii) For Block #412-4-7 in Medellín, the total number of eligible persons across all participating households is 13. For this Block, the total number of eligible persons across the participating households where a selected person participates is 8. Therefore,

$$A_{bh} = \sum_{h=1}^{n_{bk}} a_{hi} = 13 ; \quad A_{bhp} = \sum_{h=1}^{n_{bk}} a_{hp} = 8$$

$$R_{bhi} = \frac{A_{bh}}{A_{bhp}} = \frac{13}{8} \approx 1.625$$

Table 15 provides the sample counts of the Final Response Category for selected persons by Stratum.

**Table 15: Sample Counts (n) by Selected Person (SP) Final Response Category for Participating Households**

<b>Stratum</b>	<b>Metro Area</b>	<b>1-Participant SP</b>	<b>2-Non-participant Eligible SP</b>	<b>Total</b>
1-Certainty MAs	1-Bogota	956	70	<b>1026</b>
	2-Medellin A.M.	425	16	<b>441</b>
	3-Cali A.M.	292	27	<b>319</b>
	4-Barranquilla A.M.	226	17	<b>243</b>
	5-Bucaramanga A.M.	138	31	<b>169</b>
<b>Sub-Total (Certainty MAs)</b>		2037	161	<b>2198</b>
2-Probabilistic MAs	7-Manizales A.M.	122	9	<b>131</b>
	9-Villavicencio	109	3	<b>112</b>
	11-Cucuta.	207	18	<b>225</b>
	13-Ibague	142	16	<b>158</b>
<b>Sub-Total (Probabilistic MAs)</b>		580	46	<b>626</b>
<b>Grand Total</b>		<b>2617</b>	<b>207</b>	<b>2824</b>

Based on the above table, there are 2617 participating selected persons for which a final non-zero person weight is included in the final weighted data file.

### 5.7.3 Final Third-stage Person Weight

The final person weight is the weight used for producing person-level estimates, such as the proportion of persons with a specific attribute. This weight is appended to the STEP data file.

The final person weight is calculated as the product of the final household weight (which incorporates the Dwelling weight, the Household weight, the Block weight, the Block Segment weight, the Section weight, and the Metropolitan Area weight), the basic seventh stage person weight, and the response adjustment for the basic person weight. Therefore, for the selected person 'i' in Household 'h' in Dwelling 'd' in Block 'b' in Section 's' in Metropolitan Area 'j', the final person weight,  $W_{SP}$  is calculated as follows:

$$39) \quad W7_{SP} = W6_H * W7_{hi} * R_{bhi}$$

➤ The final third stage person weight is appended to the STEP data file.

#### **Example 18: Third Stage – Final Person Weight**

- i) For Block #201-2-3 in Medellín, selected household # 1 with 2 eligible persons, the selected person was a participant. All selected persons participated in For Block #201-2-3 in Medellín, therefore,  $R_{jk} = 1$

$$W7_{SP} = 1603.118 * 2 * 1 \approx 3206.24$$

- ii) For Block #412-4-7 in Medellín, in participant household # 19 with 4 eligible persons, the selected person was a participant. For this household,

$$W6_H \approx 1748.315; \quad W7_{hi} = 2; \quad R_{bhi} \approx 1.625$$

$$W_{SP} = 1748.315 * 2 * 1.625 \approx 5682.02$$

The final person weight is similarly calculated for all sampled persons. Note that the final person weight will be zero for all non-participating selected persons.

### 5.7.3.1 Weighted Person Counts

The weighted count of the non-institutionalized persons 15 to 64 years of age (inclusive) living in private dwellings in urban areas of the 13 Metropolitan Areas of Colombia at the time of data collection is provided in Table 16. The 'Weighted # of Eligible Persons' in this table is the estimated number of persons aged 15 to 64 in the Colombia target population based on the STEP sample.

**Table 16: Number of Participating Selected Persons & Weighted Total**

Stratum	Metro Area	1-Participant SP	Weighted # of Persons aged 15-64	Total
1-Certainty MAs	1-Bogota	956	4,830,295	<b>1026</b>
	2-Medellin A.M.	425	2,069,743	<b>441</b>
	3-Cali A.M.	292	1,455,241	<b>319</b>
	4-Barranquilla A.M.	226	1,107,025	<b>243</b>
	5-Bucaramanga A.M.	138	567,918	<b>169</b>
<b>Sub-Total (Certainty MAs)</b>		2037	10,030,222	<b>2198</b>
<b>2-Probabilistic MAs</b>		580	2,929,337	<b>626</b>
<b>Grand Total</b>		<b>2617</b>	<b>12,959,559</b>	<b>2824</b>



## 6 Benchmark Weights – Adjustment of Weights to Known Population Totals

The Colombia team provided known population totals to adjust the relevant STEP survey weights at the Dwelling Level and at the Person level.

### 6.1 Benchmark Adjusted Weight – Dwelling Level

Annex 1, shown below, of the document '*FINAL SAMPLING PLAN (ARD-397)-June 2013*', i.e., the Word document '[ARD\\_397\\_Final\\_sampling\\_plan\\_Colombia\\_060813](#)', provides the most current known number of Dwellings by Socioeconomic Stratum and by Metropolitan Area.

**Table 17:**

**ANNEX 1**

#### **Adjustment variable at the dwelling level**

Residential units per socioeconomic stratum, according to electric utility bill							
City or metropolitan area	Stratum 1	Stratum 2	Stratum 3	Stratum 4	Stratum 5	Stratum 6	Total
A.M MEDELLIN	88,316	301,741	292,988	86,467	62,239	29,173	860,924
A.M B/QUILLA	81,582	76,292	55,512	22,627	12,371	8,832	257,216
BOGOTA	111,591	550,211	591,398	200,214	73,199	57,723	1,584,336
CARTAGENA	50,145	51,093	32,256	10,959	6,755	7,180	158,388
A.M MANIZALES	9,798	25,515	41,308	13,984	4,557	6,927	102,089
MONTERIA	28,016	18,055	7,546	2,364	1,618	898	58,497
VILLAVICENCIO	15,554	29,640	38,592	6,764	2,592	917	94,059
PASTO	12,524	26,079	20,403	6,541	2,209	142	67,898
A.M CUCUTA	40,220	80,163	39,859	15,291	2,850	187	178,570
A.M PEREIRA	16,994	47,979	47,362	18,717	9,465	6,558	147,075
A.M B/MANGA	24,178	61,791	71,752	44,802	6,195	6,941	215,659
IBAGUE	18,443	51,529	29,732	10,772	1,675	628	112,779
A.M CALI	79,787	131,629	163,602	49,228	43,507	12,525	480,278
<b>Total</b>	<b>577,148</b>	<b>1,451,717</b>	<b>1,432,310</b>	<b>488,730</b>	<b>229,232</b>	<b>138,631</b>	<b>4,317,768</b>

Source: Public Utilities Superintendent

For the purpose of adjusting the final design weights at the dwelling level the following benchmark groups and corresponding totals were derived from the above Annex1:

<b>Table 18: Benchmark Groups and Dwelling Counts (from Annex 1 &amp; Colombia Weighted Data File)</b>			
<b>Survey Stratum (i.e., estr_mu)</b>	<b>Benchmark Group</b>	<b>Total Dwellings (Annex 1)</b>	<b>Total Dwellings COL Data File</b>
1	2-Medellin A.M.	860,924	727,998
2	A.M B/QUILLA	257,216	354,492
3	1-Bogota	1,584,336	1,644,480
4	5-Bucaramanga A.M.	215,659	225,930
5	3-Cali A.M.	480,278	473, 542
6	7-Manizales A.M., 9-Villavicencio, 11-Cucuta, 13-Ibague	919,355	1,103,090
<b>Total</b>		<b>4,317,768</b>	<b>4,529,532</b>

The final dwelling weight, adjusted for the known benchmark dwelling totals in Annex 1, i.e.,  $W_{D,BM}$ , is calculated as follows:

$$40) \quad W_{D,BM} = W_D * \frac{N_G}{\hat{N}_G}$$

where,

$N_G$  - the total number of dwellings in Benchmark Group 'G', based on Annex 1,

$\hat{N}_G$  - the total estimated number of dwellings in Benchmark Group 'G', based on the Colombia final data file.

## 6.2 Benchmark Adjusted Weight – Person Level

Annex 2, shown below, of the document '*FINAL SAMPLING PLAN (ARD-397)-June 2013*', provides the most current known number of Persons aged 15-64 by Gender and by Metropolitan Area.

Table 19:

### ANNEX 2

Adjustment variable at the person level

2012 projections for urban population aged 15 to 64 by city or metropolitan area and sex

City or metropolitan area	Urban population aged 15-64	Urban males aged 15-64	Urban females aged 15-64
A.M MEDELLIN	2,430,680	1,148,982	1,281,698
A.M B/QUILLA	1,178,441	571,757	606,684
BOGOTA	5,237,117	2,508,532	2,728,586
A.M B/MANGA	723,670	347,006	376,663
A.M CALI	656,133	315,263	340,870
CARTAGENA	612,215	293,167	319,048
A.M MANIZALES	283,100	134,297	148,803
MONTERIA	211,624	101,362	110,261
VILLAVICENCIO	287,293	137,034	150,259
PASTO	239,170	114,330	124,840
A.M CUCUTA	504,950	242,451	262,499
A.M PEREIRA	411,478	195,308	216,170
IBAGUE	332,608	159,814	172,794
<b>Total</b>	<b>13,108,479</b>	<b>6,269,303</b>	<b>6,839,175</b>

For the purpose of adjusting the final design weights at the person level the following benchmark groups and corresponding totals were derived from the above Annex2:

<b>Table 20: Person Counts by Benchmark Group &amp; Gender (from Annex 1)</b>				
<b>Survey Stratum</b> <b>(i.e., estr_mu)</b>	<b>Benchmark Group</b>	<b>Urban population</b> <b>aged 15-64</b>	<b>Urban males</b> <b>aged 15-64</b>	<b>Urban females</b> <b>aged 15-64</b>
1	2-Medellin A.M.	2,430,680	1,148,982	1,281,698
2	A.M B/QUILLA	1,178,441	571,757	606,684
3	1-Bogota	5,237,117	2,508,532	2,728,586
4	5-Bucaramanga A.M.	723,670	347,006	376,663
5	3-Cali A.M.	656,133	315,263	340,870
6	7-Manizales A.M., 9-Villavicencio, 11-Cucuta, 13-Ibague	612,215	293,167	319,048
<b>Total</b>		<b>13,108,479</b>	<b>6,269,303</b>	<b>6,839,175</b>

Table 21 provides the corresponding table of person counts from the Colombia weighted data file.

<b>Table 21: Person Counts by Benchmark Group &amp; Gender (from Colombia Weighted Data File)</b>				
<b>Survey Stratum</b> <b>(i.e., estr_mu)</b>	<b>Benchmark Group</b>	<b>Urban population</b> <b>aged 15-64</b>	<b>Urban males</b> <b>aged 15-64</b>	<b>Urban females</b> <b>aged 15-64</b>
1	2-Medellin A.M.	2,069,743	969,752	1,099,991
2	A.M B/QUILLA	1,107,025	472,608	634,417
3	1-Bogota	4,830,295	1,794,044	3,036,252
4	5-Bucaramanga A.M.	567,918	216,472	351,446
5	3-Cali A.M.	1,455,241	535,722	919,519
6	7-Manizales A.M., 9-Villavicencio, 11-Cucuta, 13-Ibague	2,929,337	1,282,021	1,647,316
<b>Total</b>		<b>12,959,559</b>	<b>5,270,618</b>	<b>7,688,941</b>

The final person weight, adjusted for the known benchmark person totals in Annex 2, i.e.,  $W_{SP,BM}$ , is calculated as follows:

$$41) \quad W_{SP,BM} = W_{SP} * \frac{N_{SG}}{\hat{N}_{SG}}$$

where,

$N_{SG}$  - the total number of dwellings in Benchmark Person Group 'SG', based on Annex 1;

$\hat{N}_{SG}$  - the total estimated number of dwellings in Benchmark Person Group 'SG', based on the Colombia final data file;

S - 1 to 6, i.e., stratum 1 to 6;

G - Male, Female.